

APPENDAGE DEVELOPMENT AND EARLY *DISTAL-LESS* REGULATION IN
ARTHROPODS: A STUDY OF THE CHELICERATE *TETRANYCHUS URTICAE*
(ACARIDA)



*The Wily Grassier by S.H. Sime (1867-1941) from
'The Sime Zoology: Beasts That Might Have Been'*

Chloë Cyrus-Kent

Corpus Christi College
University Museum of Zoology Cambridge

Supervisor: Prof Michael E. Akam FRS
Co-supervisor: Dr Max J. Telford

Thesis submitted for the degree of Doctor of Philosophy

Easter 2007

To Katie,

who has been disabled by ME throughout the duration of my PhD, but whose friendship has remained a blessing and an inspiration. Here is gratitude, awe and faith in your future.

www.afme.org.uk

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. This dissertation does not exceed 300 single-sided pages of text, excluding References and Appendices.

Chloë Cyrus-Kent

April 2007

ACKNOWLEDGEMENTS

'Our aim is to recognise each other and for each one of us to learn to see and honour in the other what he is – the counterpart and complement of the other.' Herman Hesse

I am aware of a multitude of people I would like to thank for their special contribution to my life and work during the more-than-anticipated number of years that I have been engrossed in my PhD. I look forward to sharing my gratitude in person soon, but here are just a few words and names that will have to suffice in the meantime!

Thanks to my supervisors Max Telford and Michael Akam for encouragement, enthusiasm and help when I needed it (and finally learned to ask). Thanks to Pat Simpson and Adrian Friday, my valued advisors in a few key moments. Heaps of thanks to all the many members of the Akam-Telford labs that I have had the pleasure of sharing an international, amusing, tasty and stimulating workplace with. Especially Sara Bourlat, Ron Jenner, Martin Jäkel, Eddy Nason, Judith Gallagher, Angela Stebbings, Jo Richardson, Kristen Panfilio, Andy Peel, Cassandra Extavour, Nicolas Gompel, Tassos Pavlopolous, Michalis Averof, Xaris Kondarakis, Ariel Chipman, Joakim Eriksson, Niko Prpic, Luis de Navas, Chuck Cook, Christen Mirth, Claudio Alonso, Fernando Roch, Élio Sucena and Peter Dearden, who first taught me to marvel at arthropod embryos. Many thanks to George Speller, Phil Cox, Esther Sharp, Roz Wade, Mat Lowe, Steve R, Viviane C, Rob Asher, Farnon Ellwood, Sarah Ferguson, Julie McArthur, Stu Turner and Russell Stebbings for making me feel at home in the Museum and putting up with me while writing up. Zoology would not have been the same without the kindness and support of Linda Wheatley, John Johnson, Linda Blades, Bill Lee, Val Tiplady and Bobby.

Thankyou to Wim Damen, Michael Schoppmeier, Nipam Patel, Matthias Geberdinger, Miodrag and Vava Grbic and Abderrahman (Rahim) Khila for generously hosting and helping me while in their labs at the Universität zu Köln, University of Chicago and University of Western Ontario. My gratitude for inspiration and lasting friendship in those labs also goes to Carlos Jaramillo, Matt Giorgianni, Angelika Stollewerk, Niko Prpic, Pegah Velazideh, Anica and Simone Kienle. Thanks to C. Donly at Agriculture Canada's Southern Crop Protection and Research Centre in Ontario for *Tetranychus* EST sequences and cDNA library; W. Damen for *Cupiennius* embryos; G. Panganiban for the *anti-Dll* antibody; C. Heldin and the Ludwig Institute in Uppsalla for *anti-phospho-Smad* antibody; Franck Simonnet for '*Vive les Chelicères*' solidarity. Thanks go to Corpus Christi College

(especially Paula McPhee), BBSRC, The Newton Trust and The Amateurs Trust for that collegiate and financial support that made this learning opportunity possible.

For making my world a wondrous place to be; for encouragement, patience, fun and deeply valued friendship I owe so much to: Ele H, Tonja S, Karen W, Sarah E, Chloe H, Sally A, Tanya M, Katherine C, Rachel James, Jennie L, Katie D, Hughcumber C M, Adrian Kwabena Jeans, Laura Keech (nearly), Claire H, Ruth G, Rachel McGill, Verena D-B, Ruth T, Sara & Chad P, Giorgos P, Lila K, Laura W, Laura Basell, Olivier F, the elusive Florian Nigsch, Binoy K, Vicky E, Birgit R, Il Principe Giuseppe S-T, Severin E, Alfredo J Casas Argüello, Laura H R, Bahram R, Julius R, Peppe at Café Rouge, Jo and Vanessa at Neal's Yard. Gabriel *Hollywood* de la Champag-nah and Will T Swaney shall ever have my gratitude for reaching me, believing in me and pointing me lightly back to myself. I thank R D 'Austin' Hedley for only ever being himself. I deeply thank all my family for their love and friendship: in particular, my hero of a brother Eliot for endless good cheer and real humanity, and my dad Alan for rescuing me in gloomy times, spoiling me with lunch, loot and discussions of molecular symmetry. Thanks to my wonderful, determined sister Ula for being there – and getting there first (darn it). To my brave, visionary and cute maman Julie, I owe more than my words can say. To precious Winnie, thanks for being the blue eyes and 'the money' - every penny. Thanks too to Alison, Jack, Lucy, Polly, Dot and Geoff for their special part...

An oceanic thank-you to Caroline Lea-Cox and Gill Chumbley for being beside me through many valuable struggles and a long illness that has shaped me and taught me so much about the joy of being *and* doing (and spiders).

Final thanks go to Laura Blackburn, without whom I would have exploded long ago and run away to the Med: I've appreciated the rocks, granola, patient advice, good taste in murder and all...

ABSTRACT

A major goal of evolutionary developmental biology is to explore mechanisms and events underlying evolution of the myriad body plan morphologies expressed both genetically and phenotypically within the animal kingdom. Arthropods exhibit an astounding array of morphological diversity both within and between representative sub-phyla, thus providing an ideal phylum through which to address questions of body plan innovation and diversification. Major arthropod groups are recognised and defined by the distinct form and number of articulated appendages present along the antero-posterior axis of their segmented bodies.

A great deal is known about the developmental genetics of limb development in the model insect *Drosophila melanogaster*, added to which, much comparative gene expression data and a growing body of functional genetic data is emerging for other arthropod species. Arthropod limb primordia are consistently marked by expression of the homeobox gene *Distal-less* (*Dll*), and the focus of this thesis is to compare signalling mediated by early *Dll* regulatory genes activity along antero-posterior and dorso-ventral embryonic axes during limb specification in *Drosophila*, with the activity of their orthologs in the widely disparate chelicerate, the spider mite *Tetranychus urticae* – interpreting new data with that available for other arthropods.

Having made a detailed study of spider mite embryonic (and post-embryonic) development, to provide a basis for understanding mRNA transcription and protein activity patterns, I confirmed typical expression of *Tetranychus Dll* in prosomal limb primordia. I obtained limited results for the candidate antero-posterior positioning genes *wingless* and *engrailed*, although one of the two *engrailed* paralogs I identified is reportedly expressed in posterior segmental compartments, consistent with possible conservation of Engrailed-Wingless interactions in metameric patterning and positive regulation of *Dll* in arthropod limb specification. In *Drosophila*, *wingless*-dependent *Dll* transcription is restricted along the dorso-ventral axis by dorsal Dpp-mediated and ventral EGFR-mediated signalling gradients. Based on data from *Tetranychus* and other arthropods, neither dorsal nor ventral signalling regimes appear conserved outside the *Drosophila* system. *Dll* suppression in fly abdominal segments occurs due to powerful Hox (Ubx/AbdA) repression of the early *Dll cis*-regulatory element; this is discussed in relation to the independently evolved limbless chelicerate opisthosoma, informed by hypothetical scenarios of *cis* (regulatory DNA) and *trans* (coding sequence) evolution.

Given practical difficulties and limitations encountered while working with spider mites, I offer a final assessment of the place of *Tetranychus urticae* as a non-model, and yet still valuable chelicerate species to consider carrying into the exciting future of evolutionary developmental biology.

CONTENTS

ACKNOWLEDGEMENTS p.4

ABSTRACT p.6

CHAPTER I: INTRODUCTION

- 1.1 Evolutionary developmental biology and the Arthropoda p.14
 - 1.1.1 The origins of evolutionary developmental biology
 - 1.1.2 Sources of developmental genetic variation?
 - 1.1.3 Arthropoda as a model phylum for studying body plan diversity
- 1.2 Arthropod appendages and *Distal-less* regulation p.21
 - 1.2.1 The *Distal-less* regulatory network in *Drosophila melanogaster*
 - 1.2.2 The *Distal-less* regulatory network in divergent arthropods: moving beyond the *Drosophila* paradigm
 - 1.2.3 An extra player? Regulation of limb specification by *Sp* genes
- 1.3 Choice of an organism: *Tetranychus urticae* p.37
 - 1.3.1 Arthropod phylogeny – pertinence of the Chelicerata
 - 1.3.2 The spider mite: a Chelicerate model for evo-devo?
- 1.4 Aims and Structure of this Thesis p.43
 - 1.4.1 Thesis Aims
 - 1.4.2 Thesis Structure

CHAPTER II: OVERVIEW OF DEVELOPMENT IN THE TWO-SPOTTED SPIDER MITE *TETRANYCHUS URTICAE* KOCH 1836

- Introduction to Chapter II p.46
- 2.1 Embryonic development p.47
 - 2.1.1 Cleavage and blastoderm formation
 - 2.1.2 Germdisc stage
 - 2.1.3 Germband formation and development
 - 2.1.4 Appendage development
- 2.2 Post-Embryonic Development p.55
 - 2.2.1 Prelarva
 - 2.2.2 Hexapod larva
 - 2.2.3 Nymphal stages
 - 2.2.4 Adult stage
 - 2.2.5 Reproduction
- Conclusions to Chapter II p.61

CHAPTER III: APPENDAGE SPECIFICATION IN *TETRANYCHUS URTICAE*: *DISTAL-LESS* AND *SP* GENES

Introduction to Chapter III	p.65
3.1 Cloning of the <i>Tetranychus urticae</i> <i>Distal-less</i> gene	p.67
3.1.1 Degenerate PCR screening	
3.1.2 Inverse PCR and EST data	
3.1.3 Phylogenetic analysis: Bayesian inference	
3.1.4 Digoxigenin-labelled mRNA probe synthesis and efficacy	
3.2 <i>Tu-Dll</i> gene expression during embryogenesis	p.70
3.2.1 Absence of <i>Tu-Dll</i> transcriptional activity (0-20hr AEL)	
3.2.2 <i>Tu-Dll</i> transcription associated with ventral limb primordia in the late germband (21-22hr AEL)	
3.2.3 <i>Tu-Dll</i> transcription in developing limb buds and pre-chelicerel CNS derivatives (23-30hr AEL)	
3.2.4 Dynamic <i>Tu-Dll</i> transcription during ventral ridge contraction (30-39hr AEL)	
3.2.5 <i>Tu-Dll</i> transcriptional modulation within appendages during ventral closure (40-45hr AEL)	
3.2.6 <i>Dll</i> gene activity in appendages and the ocular segment during germband contraction and dorsal closure (45-50hrs AEL)	
3.3 <i>Tetranychus</i> Distal-less protein: antibody tests	p.79
The polyclonal Dll antibody	
Lack of antibody reactivity in <i>Tetranychus</i> embryos	
3.4 Cloning of <i>Tetranychus</i> <i>Sp8/9</i> and <i>Sp1/5/4</i> orthologs	p.81
3.4.1 Degenerate PCR screening	
3.4.2 Phylogenetic analysis	
3.4.3 Note on <i>Tetranychus</i> <i>Sp</i> gene expression	
Conclusions to Chapter III	p.84

CHAPTER IV: APPENDAGE SPECIFICATION IN *TETRANYCHUS URTICAE*: *WINGLESS*, *WNT* GENES AND *ENGRAILED*

Introduction to Chapter IV	p.87
4.1 Cloning <i>Tetranychus urticae</i> <i>Wnt</i> genes	p.89
4.1.1 Degenerate PCR screen for a <i>Tetranychus</i> <i>Wg/Wnt1</i> ortholog	
4.1.2 <i>Tetranychus</i> <i>Wnt</i> genes from an EST screen	
4.1.3 Phylogenetic analysis of <i>Tu-Wnt</i> genes	
4.2 Detecting Wnt1 protein activity in <i>Tetranychus urticae</i> with Wg and Armadillo antibodies	p.93
4.2.1 The Wnt signal pathway: antibody targets	
4.2.2 Wingless monoclonal antibody 4D4	
4.2.3 Armadillo/ β -Catenin antibodies	
4.2.4 Conserved arthropod Wnt1 roles: <i>Tetranychus</i> predictions?	

4.3 Cloning <i>Tetranychus engrailed</i> (<i>Tu-en</i>) genes	p.97
4.3.1 Degenerate PCR and sequence analysis: Two <i>Tu-en</i> genes	
4.3.2 Inverse PCR (iPCR): Further <i>en-1</i> and <i>en-2</i> sequence data	
4.3.3 Phylogenetic analysis of <i>Tu-en1</i> and <i>Tu-en2</i>	
4.3.4 <i>Tu-en</i> ssRNA probe synthesis	
4.4 <i>Tetranychus engrailed</i> mRNA expression	p.103
4.4.1 <i>Tu-en</i> mRNA transcription in whole mount embryos	
4.4.2 Successful detection of <i>Tu-en</i> mRNA transcripts by others	
4.5 <i>Tetranychus</i> Engrailed protein expression	p.106
4.5.1 Monoclonal Engrailed and Inverted antibodies 4F11 and 4D9	
4.5.2 Polyclonal Engrailed antibody 'α-ht-en'	
Conclusions to Chapter IV	p.108

CHAPTER V: APPENDAGE SPECIFICATION BY CANDIDATE DORSO-VENTRAL PATTERNING GENES: *TETRANYCHUS DPP* AND *EGFR*

Introduction to Chapter V	p.112
5.1 Cloning <i>Tetranychus decapentaplegic</i>	p.114
5.1.1 Degenerate PCR screening for <i>dpp</i>	
5.1.2 Further <i>Tu-dpp</i> data from inverse PCR and cDNA library screening	
5.1.3 <i>Tu-dpp</i> ssRNA probe synthesis	
5.2 <i>Tu-dpp</i> mRNA expression during embryogenesis	p.118
5.2.1 <i>Tu-dpp</i> mRNA detection: negative results	
5.2.2 Reasons for negative results	
5.3 Detection of <i>dpp</i> signal transduction in <i>Tetranychus urticae</i> and <i>Cupiennius salei</i> embryos <i>via</i> a phospho-Smad-1 antibody	p.119
5.3.1 Rationale for detecting Dpp activity <i>via</i> phospho-Smad-1	
5.3.2 <i>Tetranychus urticae</i> p-Smad-1 distribution	
5.3.3 <i>Cupiennius salei</i> p-Smad-1 distribution	
5.3.4 Smad-1 phosphorylation: <i>Cupiennius</i> vs. <i>Tetranychus</i>	
5.4 Cloning <i>Tetranychus urticae</i> EGF receptor (<i>TER</i>)	p.127
5.4.1 Two putative <i>Tetranychus</i> <i>EGFR</i> genes from degenerate PCR	
5.4.2 Extension of <i>TER-a</i> sequence data by cDNA library screening and inverse PCR methods	
5.4.3 <i>TER-a</i> DIG-labelled ssRNA probes	
5.4.4 Attempts to extend <i>TER-b</i> sequence data by cDNA library screening and inverse PCR methods	
5.4.5 Critical assessment of <i>TER-b</i> authenticity	
5.5 <i>TER-a</i> mRNA transcription during <i>Tetranychus</i> embryogenesis	p.134
5.5.1 Blastoderm - germdisc stage: no <i>TER-a</i> transcripts	
5.5.2 <i>TER-a</i> at germband stage: onset of transcription	
5.5.3 Early limb bud stage: ventral neurogenic <i>TER-a</i> transcripts	
5.5.4 <i>TER-a</i> transcription during late limb bud development	
5.5.5 <i>TER-a</i> transcription during ventral ridge contraction/limb elongation	
5.5.6 <i>TER-a</i> transcription during ventral closure	
5.5.7 <i>TER-a</i> transcription during germband contraction and dorsal closure	

5.6 Testing for <i>EGF</i> signalling in <i>Tetranychus</i> embryogenesis with a dpMAPK antibody	p.143
5.6.1 Rationale for using a dp-MAPK antibody	
5.6.2 Lack of <i>Tetranychus</i> cross-reactivity with anti-dp-MAPK	
Conclusions to Chapter V	p.145

CHAPTER VI: APPENDAGE SPECIFICATION BY CANDIDATE ANTERO-POSTERIOR PATTERNING GENES: *TETRANYCHUS UBX* AND *ABD-A*

Introduction to Chapter VI	p.149
6.1 Cloning and analysis of <i>Tetranychus Hox</i> genes	p.151
6.1.1 Amplification of <i>Tetranychus Hox</i> orthologs by degenerate PCR	
6.1.2 <i>Tu-Ubx</i> sequence extension and analysis	
6.1.3 <i>Tu-Ubx</i> single-stranded RNA probe synthesis	
6.1.4 Absence of <i>Tu-abdA</i> from the <i>Tetranychus urticae</i> genome	
6.2 <i>Tu-Ubx</i> mRNA transcription during embryogenesis	p.158
6.2.1 Lack of detectable <i>Tu-Ubx</i> transcripts: 0 – 17hr AEL	
6.2.2 Early <i>Ubx</i> expression	
6.2.3 Later <i>Ubx</i> expression	
6.2.4 Inferences and hypotheses on <i>Tetranychus Ubx/Dll</i> interactions	
6.3 Testing for <i>Ubx/abdA</i> protein activity in <i>Tetranychus</i> with antibody FP6.87 (α -UbdA)	p.162
6.3.1 The UbdA/FP6.87 antibody	
6.3.2 Lack of anti-UbdA cross-reactivity in <i>Tetranychus</i>	
6.3.3 Assessment of negative results based on epitope conservation	
Conclusions to Chapter VI	p.165

CHAPTER VII: DISCUSSION

7.1 Assessment of conservation within the <i>Dll</i> genetic regulatory network for early limb development	p.168
7.1.1 Antero-posterior specification of limb primordia	
7.1.2 Dorso-ventral specification of limb primordia	
7.1.3 Deep genetic homology in limb development?	
7.2 The limits of comparative gene expression studies	p.173
7.2.1 Example A: Affinity of the arthropod labrum	
7.2.2 Example B: Sub-functionalisation among <i>engrailed</i> paralogs	
7.2.3 Example C: <i>Dll</i> repression in the chelicerate opisthosoma	
7.2.4 Example D: Variable interactions and dissociation between segment polarity gene expression and function in limb specification	
7.2.5 Verdict on genetic inferences based on <i>Tetranychus urticae</i> data	
7.3 Assessment of <i>Tetranychus urticae</i> as a model chelicerate	p.176
7.3.1 Critique of practical workability	
7.3.2 Assessment of progress on genomics and gene function	
7.3.3 Verdict on <i>Tetranychus</i> as a model organism for evo-devo	

7.4 Potential future directions for the study of spider mite limb development and evo-devo p.180

7.4.1 Loose ends in this study of *Tetranychus Dll* regulation

7.4.2 Last word: The place of *Tetranychus* in the future of evo-devo

CHAPTER VIII: MATERIALS AND METHODS

8.1 *Tetranychus urticae* embryo preparation p.184

8.1.1 Spider mite husbandry

8.1.2 Spider mite embryo collection

8.1.3 Fixation of spider mite embryos for *in situ* hybridisation

8.1.4 Fixation of spider mite embryos for antibody staining

8.2 Molecular Cloning and Sequencing p.187

8.2.1 Degenerate PCR

8.2.2 Reverse Transcriptase PCR

8.2.3 Inverse PCR

8.2.4 cDNA library screening

8.2.5 Cloning

8.2.6 Gene sequencing

8.3 Phylogenetic analysis of putative orthologs p.192

8.3.1 Molecular phylogenetics theory

8.3.2 Preliminary gene identification

8.3.3 Generation of multiple nucleotide/amino acid alignments

8.3.4 Defining Exclusion and Inclusion Sets

8.3.5 Bayesian analysis

8.3.6 Maximum Likelihood analysis by Quartet Puzzling

8.3.7 Comparative protein sequence analysis

8.4 *In situ* hybridization p.201

8.4.1 DIG labeled mRNA probe synthesis

8.4.2 mRNA hybridization to test probe labeling efficiency

8.4.3 Whole-mount mRNA *in situ* hybridization protocol

8.4.4 DAPI nuclear counter-staining

8.5 Immuno-histochemical staining p.209

8.5.1 Antibody staining in *Tetranychus urticae*

8.5.2 Antibody staining in *Cupiennius salei*

8.5.3 Primary and Secondary antibodies

8.6 Microscopy and Imaging p.214

8.6.1 Mounting stained embryos

8.6.2 Photomicrography

8.6.3 Scanning electron microscopy

8.6.4 Image processing

REFERENCES p.218

APPENDICES

Appendix 1: Spider mite chaetotaxy	p.239
Appendix 2: Oligonucleotide primers	p.240
2.1 <i>Distal-less</i> primers	
2.2 <i>Sp</i> primers	
2.3 <i>Wingless/Wnt</i> primers	
2.4 <i>Engrailed</i> primers	
2.5 <i>Decapentaplegic</i> primers	
2.6 <i>Epidermal growth factor receptor</i> primers	
2.7 <i>Hox</i> gene primers	
Appendix 3: Non-target gene sequences and analysis	p.245
3.1 <i>Tetranychus</i> receptor tyrosine kinase (RTK) <i>Src</i>	
3.2 A <i>Tetranychus</i> <i>Tho</i> complex (<i>Tho-C</i>) gene	
3.3 <i>Tetranychus</i> <i>Sex combs reduced</i> (<i>Scr</i>)	
Appendix 4: Notes on phylogenetic analysis protocols	p.248
4.1 TranslatorX3 commands	
4.2 Protocol for defining Exclusion and Inclusion sets	
4.3 Bayesian analysis protocol and commands	
Appendix 5: <i>Engrailed</i> gene multi-functionality	p.251
5.1 <i>Engrailed</i> and segment boundary formation	
5.2 <i>Engrailed</i> and segmental limb outgrowth	
5.3 <i>Engrailed</i> orthologs in neuro-, myo- and morpho-genesis	
5.4 Ancestral functions of arthropod <i>Engrailed</i> orthologs?	
Appendix 6: Detection of <i>Tu-Sp8/9</i> and <i>Tu-Sp1/3/4</i> transcripts with short ssRNA probes	p.254
6.1 Synthesis of <i>Tu-Sp1/3/4</i> and <i>Sp8/9</i> RNA probes	
6.2 Expression of <i>Tu-Sp1/3/4</i> and <i>Sp8/9</i> during embryogenesis	
Appendix 7: Amended interpretation of data published on early segmentation and germ cell fate specification in <i>Tetranychus urticae</i>	p.257
7.1 Early segmental patterning	
7.2 Germ line specification	

CHAPTER I

INTRODUCTION



Scorpion d'Europe

1.1 Evolutionary developmental biology and the Arthropoda

The scientific field of evolutionary developmental biology (abbreviated to 'evo-devo') aims to comprehend the mechanisms and events that have generated the incredible diversity in animal and plant forms that we observe today. Within this broad aim, a major goal, and the focus of this thesis, is in exploring mechanisms and events underlying evolution of the myriad body plan morphologies expressed both genetically and phenotypically within the Kingdom Animalia. The arthropods exhibit an astounding array of morphological diversity both within and between all representative sub-phyla, providing an ideal phylum through which to address questions of body plan innovation and diversification.

1.1.1 The origins of evolutionary developmental biology

Recognition of evo-devo as a coherent field occurred in the 1980s, with an awareness that it was in fact more of a re-birth than a new discipline: interest in explaining and rationalising the richness of extant and extinct animal form has preoccupied thinkers from Greek philosophers to modern-day biologists (Clark 2000; Gilbert 2000). Animal development involves the unfolding of complex genetic programs in space and time to build phenotypes diverse in form and function that, within the constraints of any developmental bias, are adapted *via* natural selection to their specific environments (Averof 1994). There is a need therefore to address both development and evolution when questioning animal diversity, but this need has not always been recognised:

Embryology and variation in form

Aristotle (4th century B.C.) compared embryo development in many species, especially mammals, observing distinct modes of zygotic cleavage, growth and increased complexity ('epigenesis') over time (Gilbert 2000; Martinez Arias and Stewart 2002). Unlike Plato's idea that human reason could uncover actual truth¹ (the philosophical basis for the kind of claims that science and Darwin later made), Aristotle relied on his fallible, 'common sense' to question variation in development and form, his conclusions centring on the notion that human form was a template for all others, by virtue of 'hybridisation' or 'deformation' of the pre-existing 'Ur-Human' pattern (Clark 2000). In the late 18th century, Goethe established the concept of homology to highlight structural unity between distinct species, although as a proponent of 'perfection' in animal design according to Platonic ideal

¹ Plato wrote "*the Intellect of Man is a spark of the light by which the world was created*", crediting reason as Science does in its search for fact, but unlike Science, assuming a theistic final cause for Nature (Clark 2000).

‘archetypes’, he did not predict change in form over time. Comparative embryology advanced in parallel with improved microscopy, so that in the early 19th century the ‘Rational Morphologists’ were divided between those (e.g. Cuvier, Bell) who focussed on inter-specific adaptations to ‘conditions of existence’ and those (e.g. St Hilaire, Owens) who considered adaptations secondary to the fundamental unity of type or design, i.e. homology (Clark 2000; Gilbert 2000). Experimental embryology was born in the late 19th century, pioneered by characters such as Weismann, Roux² and Driesch who studied the role of nuclear factors in determinative (mollusc) vs. regulative (sea urchin) embryogenesis, and Wilson and Lillie whose work on cleavage and cell lineages in diverse protostomes fuelled contrasting views, again emphasising either patterns in homology pointing to relationships (phylogenetics) or secondary adaptation (Gilbert 2000; Martinez Arias and Stewart 2002; Wilkins 2002). Darwin’s pivotal thesis, *The Origin of Species* (1859), came to offer a resolution in suggesting that embryonic homology arises from shared descent, and that adaptations arise by natural selection on variants within a given population (Darwin 1859).

Embryology without evolution, evolution without embryology

Unfortunately evolutionary theory, dominantly influenced by Darwinian natural selection, became separated from experimental embryology, and the influence of development continued to be disregarded whereas population genetics, palaeontology and systematics were later integrated in the neo-Darwinian movement (Carroll 2005). According to Wilkins (2002), this separation first occurred around 1870 when Haeckel’s ‘biogenetic law’ was proved wrong, making evo-devo thinking unpopular: Haeckel proposed that embryos pass through forms of their ancestors during ontogeny, contrary for example to von Baer who showed that developing embryos depart increasingly from features general to their phylum, acquiring specialised features without passing through stages resembling ‘lower’ animals (Gilbert 2000; Wilkins 2002).

Re-uniting embryo development and animal evolution

The mid 20th century held promise of reconciliation, exemplified by G.R. de Beer whose *Embryos and Ancestors* (1940) is prefaced with the statement that “a much better synthesis could be made of our knowledge of embryonic development and evolutionary descent, opening up new fields for observation and co-ordination of studies in embryology, genetics and evolution” (de Beer 1940). Notable potential to reverse the separation appeared in the early 1900s in the rigorous attentions of T.H. Morgan on genetics of the fruit fly *Drosophila melanogaster*, which established a powerful resource for future developmental genetics - although Morgan himself became more concerned with genetic transmission mechanisms (Martinez Arias and Stewart 2002). Although neo-Darwinism was

² Albert Roux is credited with explicitly proclaiming the benefit of an ‘evo-devo’-type approach, connecting embryology to evolutionary scenarios (Gilbert, S. F. 2000).

initially constrained by conceptualising mutations as only having extremely minimal effect, systematic mutant screens in *Drosophila* in the 1970s expanded this view by revealing numerous regulatory genes, and ‘realisator’ genes that affect large-scale pattern formation and morphogenesis (Lewis 1963). Later addition of the nematode *Caenorhabditis elegans* and chordate *Mus musculus* as intensively studied ‘model’ species furnished developmental genetics with a broader phylogenetic base, which then continued to expand to include species from multiple animal phyla and many arthropod and chordate sub-groups (Wilkins 2002). Aided by advances in molecular and morphological phylogenetics, a good degree of consensus about animal kingdom relationships also emerged during the late 20th century³, allowing the new raw material (i.e. developmental genetic data) of comparative developmental biology to be put into a phylogenetic context, informative for reconstructing evolutionary scenarios (Adoutte et al. 2000; Carroll et al. 2001; Cracraft and Donoghue 2004).

Genetic unity in animal development

Once molecular techniques allowed determination of actual gene and protein sequences it became clear that Bilaterians share a fundamental genetic unity (Carroll et al. 2001; Davidson 2001; Davidson and Erwin 2006; Wilkins 2002). A common repertoire, or ‘toolkit’ of genes and gene families was identified, including key regulatory genes encoding transcription factors (firstly and most famously the *Hox* gene cluster), as well as genes encoding intercellular signalling components, cyto-skeletal proteins, metabolic enzymes and molecules guiding conserved forms of terminal cell differentiation (Kamm et al. 2006). This discovery of genetic unity in spite of morphological diversity was a great surprise, as elucidation of the structure and function of DNA, RNA, and protein had led to a belief that animal diversity would eventually be linked to deployment of gene and protein complements unique to each distinct taxon. Furthermore, this unity highlighted an undeniable connection between embryo development and animal evolution: ‘evo-devo’ was born again. The historical origins of evolutionary developmental biology are reviewed in more detail in Martinez-Arias & Stewart (2002), Wilkins (2002) and Gilbert (2000), for example (Gilbert 2000; Martinez-Arias and Stewart 2002; Wilkins 2002).

1.1.2 Sources of developmental genetic variation?

Faced with the ‘paradox’ of a widely conserved genetic toolkit, the present challenge of evo-devo lies in explaining how diversity is generated in spite of it. It has been proposed that “evolutionarily relevant mutations” are primarily those involving *cis*-regulatory enhancer elements that control gene

³ Consensus rests on monophyly of the Bilateria, and division of the animal kingdom into three large branches: one deuterostome branch and two protostomes branches, Ecdysozoa and Lophotrochozoa (Adoutte 1999, Adoutte 2000, Carroll 2001).

transcription, but functional changes in *trans*, in sequences coding for transcription factor proteins, can also be important in shaping morphological evolution (King and Wilson 1975; Wray 2007). In addition to *cis* and *trans* elements of regulation, a compelling body of data and theory proposes that alternative regulatory levels (ARLs) may be important in shaping developmental evolution. ARLs such as regulation of alternative splicing, micro-RNAs, mRNA localisation and chromatin modelling are reviewed in Alonso & Wilkins (2005) (Alonso and Wilkins 2005). *Cis*-regulatory elements, coding sequences and trans-regulatory apparatuses are conceptualised as interacting in genetic regulatory networks (GRNs), with each node represented by a single transcription factor (TF) with its own *cis*-regulatory control element(s). It is the connections between nodes, described as GRN ‘architecture’ that are considered of principal importance to body plan evolution, as GRNs affect all levels of developmental processes (specification, patterning, terminal differentiation) and hence both small- and large-scale morphology (Davidson 2001; Davidson and Erwin 2006; Davidson et al. 2003; Wilkins 2002).

Evolution ‘in trans’

Trans-regulatory apparatus can be defined as genes which encode regulatory proteins (transcription factors) that “display high specificity for a particular *cis*-regulatory DNA sequence, and which performs some function that affects transcriptional output” (Davidson 2001). Mutations in protein coding sequence can affect post-transcriptional or post-translational processing, altering transcription factor sequence and, potentially, function. A single transcription factor may affect few or more often hundreds of downstream target genes (e.g. *Drosophila* AbdB in posterior spiracle vs. Dfd in maxilla formation), and so protein evolution is likely to have many pleiotropic phenotypic effects; such pleiotropy may create stronger negative selection relative to changes in *cis* (Hueber et al. 2007; Lohmann 2006; Wray 2007). However, gene duplication hypothetically increases available genetic resources, allowing for recruitment of modified proteins to novel functions while retaining ancestral GRN integrity *via* stabilising selection on an original paralog (Wilkins 2002). Furthermore, it has been shown that some proteins (e.g. AbdA) have a modular structure, and distinct target gene subsets respond to interaction with specific modules, such that localised sequence change has constrained potential downstream effects (Merabet et al. 2003).

Examples of *trans* evolution affecting body plan morphology include:

- a. Ultrabithorax (Hox protein) C’ modification in higher insects, repressing *Distal-less* regulatory element *Dll-504* and as a consequence, abdominal limb development (Galant and Carroll 2002; Levine 2002; Pavlopoulos and Averof 2002; Ronshaugen et al. 2002).

- b. Antennapedia N' modification in crustacean *Daphnia*, mediating novel repression of *Dll-504* and micro-evolutionary change in ventral T1 limb form (Deutsch and Mouchel-Vielh 2003; Shiga et al. 2002).
- c. In the hemipteran *Oncopeltus fasciatus*, posterior prevalence among Bithorax complex (BX-C) genes occurs by post-translational modification (rather than post-transcriptional control as in *Drosophila melanogaster*) (Angelini et al. 2005).
- d. Differences in limb and skeletal morphology among more than 90 breeds of dog (*Canis*) have been directly linked to variation in the number and length of tandem repeats within coding sequences of genes (e.g. *Aristal-less-like-4*, *Runt-related-2*) for several developmentally significant transcription factors (Fondon and Garner 2004).

Evolution 'in cis'

Cis-regulatory apparatus refers to “regions in the vicinity of each gene which contain the specific sequence motifs at which those regulatory proteins which affect its expression bind” (Davidson 2001). Evolution of *cis*-regulatory elements has been explored experimentally and by modelling, with various changes in genomic regulatory network architecture implicated as causal in small and large-scale morphological evolution: see Wray (2007) for a recent review (Davidson 2001; Wray 2007). The fundamental consequence of *cis*-regulatory evolution is altered gene expression (spatial and/or temporal), caused by changes in upstream regulation or post-transcriptional processing (Wray 2007). Point mutation or recombinatorial shuffling in *cis*-control DNA occurs such that binding sites for specific transcription factors are either gained (gene capture) or lost, and gene capture may lead to co-option or recruitment for new functions, as target genes are expressed in new territories. Recruitment may be aided by duplication events due to increased sequence available for ‘capture’ (as for protein evolution), and if paralogs act redundantly - or share an overlapping function with an unrelated gene – selection may allow sequence divergence towards new molecular functions (Laney and Biggin 1996; Wilkins 2002).

Some examples in which ‘upstream’ *cis*-regulatory divergence currently appears the most likely explanation for affecting divergent TF expression are:

- a. Among a range of crustacean Orders, shifts in anterior Ubx domains correlate with limb transformation from thoracopod (Ubx present) to maxilliped (Ubx absent) identity (Akam 2000; Averof and Patel 1997).
- b. Within the Echinodermata, homeobox genes *Dll*, *engrailed* and *orthodenticle* have been recruited to radically divergent expression domains, executing morphogenesis of the unusual echinoderm body plan (Lowe and Wray 1997).

- c. Within the vertebrates, relative modification of Hox expression domains along the antero-posterior axis is implicated in differences in axial patterning and limb identity in chick vs. mouse, and in limblessness of the snake trunk (Burke et al. 1995; Cohn and Tickle 1999).
- d. In *Drosophila*, anteriorly-expressed Hox cluster genes contain binding sites for repression by more posterior ones, generating ‘posterior prevalence’ by transcriptional regulation (Appel and Sakonju 1993).

Select examples where ‘downstream’ *cis*-regulatory evolution may be causal in affecting co-option of distinct target genes or gene batteries are:

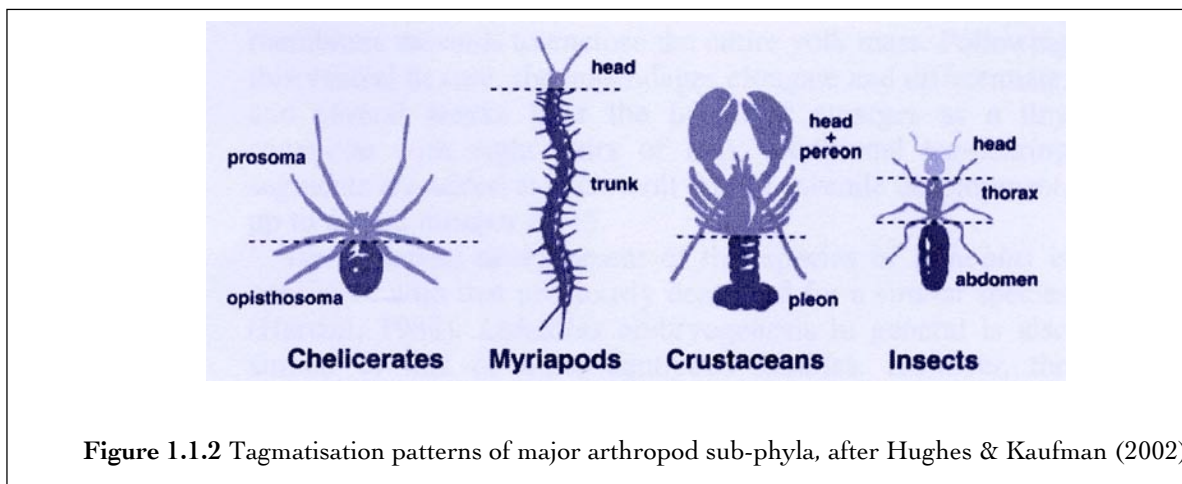
- a. After divergence of Collembola from true insect lineages, *Dll* acquired binding sites responsive to repression by AbdA, partly explaining the typical limbless state of the True insect abdomen (Angelini and Kaufman 2005b; Gebelein et al. 2002; Gebelein et al. 2004; Palopoli and Patel 1998; Vachon et al. 1992).
- b. In *Tribolium castaneum*, T2 elytron morphology depends on novel execution of ‘elytron program’ genes. ‘Elytron program’ genes are sensitive to repression by Ubx in T3, generating default to a membranous hindwing identity that is modified somewhat in form (as in Lepidoptera and Hymenoptera) by concurrent sensitivity of ancestral ‘wing program’ genes to Ubx control (Tomoyasu et al. 2005).
- c. In *Drosophila*, binding sites for Ubx repression occur in promoter elements of ‘wing’ genes (e.g. *Spalt*), so that wings can form in T2 but are suppressed in the T3 segment (where Ubx is expressed), mediating development of the T3 haltere (Galant et al. 2002; Levine 2002).

1.1.3 Arthropoda as a model phylum for studying body plan diversity

The arthropods comprise an estimated 90% of living animal species, and they display remarkable variation in form both within and between the major extant sub-phyla: Chelicerata, Myriapoda, Crustacea and Hexapoda (Moore 2001). The arthropods provide an ideal model system for studying the developmental basis of body plan evolution, as phylum Arthropoda and individual sub-phyla within it are all monophyletic groups, allowing inferences about evolutionary trends based on relative similarities and differences (Blaxter 2001; Davidson 2001; Hwang et al. 2001; Telford and Thomas 1995). Definitive arthropod features include segmentation along the body axis, a chitinous exoskeleton and jointed, articulated appendages; further shared, ‘synapomorphic’ features are documented in Brusca & Brusca (1990, 2003) (Anderson 1999; Brusca and Brusca 2003; Brusca and Brusca 1990). Phylogenetic analyses show that features such as the jointed, segmented limbs are clearly homologous, representing shared derived characters present since divergence of stem group arthropods from a common ancestor (Arthur et al. 1999; Ewing 1928; Minelli 2002). Although the

exact nature of the last common arthropod ancestor is unknown, close living outgroup taxa (onychophora, tardigrades) and a number of enigmatic fossils (e.g. *Kerygmachella*, *Anomalocaris*) point to an onychophoran-like animal with a fairly homonomous trunk, lobopodian, unbranched ('uniramous') legs and anterior sensory appendages (Akam 2000; Blaxter 2001; Budd 2001; Budd 2002; Cracraft and Donoghue 2004; Garey et al. 1996; Giribet et al. 1996; Giribet et al. 1999).

Further to a suite of true 'Euarthropod' features, each sub-phylum is characterised by a distinct tagmatisation pattern - specific morphological and functional division of the body – providing a rich means by which to address mechanisms of morphological innovation. The insect body plan is characterised by a head, thorax and limbless abdomen; the definitive crustacean tagma is the head (two antennae, mandible, two maxillae), followed by variable numbers of pereon and pleon segments; myriapod tagma include a head and elongate, limbed trunk; and chelicerates are distinguished by an anterior prosoma with gnathal and walking appendages and a limbless posterior opisthosoma⁴ (see Figure 1.1.2 below) (Barnes et al. 2001; Hughes and Kaufman 2002a; Hughes and Kaufman 2002b).



Due to extensive classical and contemporary genetic studies of *Drosophila melanogaster*, much is known about fly genetics regarding developmental processes typical to arthropods, such as segmentation and limb development (Martinez Arias and Stewart 2002; Simcox et al. 1989). As mentioned briefly before, a range of other arthropod species have been studied and compared to the model *Drosophila* paradigm, aiming to reveal evolutionary patterns of conservation and divergence in orthologous gene and genetic network deployment in shared developmental contexts (termed **comparative evo-devo**). A growing body of comparative molecular data includes gene and protein sequence and expression

⁴ The chelicerate opisthosoma technically bears modified or vestigial limbs in some of the more anterior segments, for example opisthosomal gills in marine Xiphosura, and in terrestrial forms book lungs, tubular tracheae and spinnerets. These structures are proposed to represent reduced or invaginated dorsal limb branches (epipods), derived from ancestral respiratory or osmoregulatory dorsal appendages. Prosomal limbs are consistently dominant in size, appear earlier in embryogenesis and derive from typical ventral appendage primordia (Anderson 1973, Damen 2002).

data for Higher and Lower insects, a number of crustaceans and a few myriapod and chelicerate species. When it comes to functional genetics, none of these species match the workability of *Drosophila melanogaster* and only a small – but rapidly growing - number (e.g. insects *Tribolium castaneum*, *Oncopeltus fasciatus*, crustacean *Parhyale hawaiiensis*) have so far proved easily amenable to manipulation of gene function by mutagenesis, transgenesis or RNA-interference (Angelini and Kaufman 2005a; Angelini et al. 2005; Beerman et al. 2001; Lewis et al. 2000; Pavlopoulos and Averof 2005). Nevertheless, resting on the foundation of work in *Drosophila*, the wealth of gene expression studies coupled with limited demonstrations of gene function have expanded our knowledge of certain aspects of arthropod developmental evolution, and brought old and new questions into sharper light. An important question among these, and the topic of this thesis, is the degree to which genetic regulatory interactions responsible for earliest specification of limb position along the antero-posterior axis, are conserved outside the *Drosophila* paradigm. To address this, I shall outline what is known in *Drosophila*, what comparative data we already have from other arthropod species, and what is yet required to give a satisfactory, complete picture.

1.2 Arthropod appendages and *Distal-less* regulation

The homeobox gene *Distal-less* (*Dll*) is of specific interest regarding appendage development, as it has ‘toolkit’ status among Bilaterian genes, functioning in the development of ectodermal outgrowths as disparate as vertebrate limb buds, sea urchin tube feet, polychaete parapodia and arthropod legs (Panganiban et al. 1997; Panganiban et al. 1994; Panganiban et al. 1995). The presence of *Dll* activity in a diverse range of animal appendages may indicate a deep genetic homology (discussed in Chapter VII) but the appendages themselves are analogous, specific form and tissue arrangements evolved in independent lineages to satisfy a common requirement for structures enabling locomotion, grasping and feeding (Shubin et al. 1997). If we consider the arthropods exclusively, however, their jointed limbs are demonstrably homologous, and hence it is feasible to imagine that genetic networks controlling *Dll* gene expression may be conserved between groups. Using the same logic, where *Dll* regulatory networks diverge between phyla we may reveal paths of molecular evolution that are linked to measurable changes in body plan. For example, further to a conserved role for *Dll* in limb specification and proximo-distal patterning throughout the arthropods, changes in *Dll* regulation have

been implicated in explaining the differences in appendage number and morphology between major sub-phyta (Averof and Akam 1993; Averof and Akam 1995; Levine 2002; Panganiban and Rubenstein 2002; Panganiban et al. 1995; Ronshaugen et al. 2002). This project aims to compare *Dll* regulation during early limb specification in the spider mite *Tetranychus urticae* with that known in the model insect *Drosophila melanogaster* (see section 1.4 regarding choice of *Tetranychus* as study organism). The aim of such a comparison, between two species from basally diverging arthropod lineages, is to broaden our awareness of *Dll* regulatory network conservation in different sub-phyta and pinpoint possible genetic changes that have been crucial in the developmental evolution of distinct chelicerate and insect body plans.

1.2.1 The *Distal-less* regulatory network in *Drosophila melanogaster*

In the *Drosophila* embryo, early *Dll* gene activation involves positive antero-posterior (A-P) and negative dorso-ventral (D-V) positional signals acting on an early *Dll-504* enhancer element (Cohen et al. 1993; Cohen et al. 1991; Cohen 1990; Cohen and Yurgens 1989; Diaz-Benjumea et al. 1994; Lecuit and Cohen 1997; Williams and Nagy 2001). These signals generate a segmentally reiterated co-ordinate system for precise, localised specification of ventro-lateral clusters of *Dll*-positive cells. Further to establishing *Dll* expression in each segment, the early *Dll* enhancer is repressed in all abdominal segments, generating the typical insect body-plan of a head, thorax with three pairs of thoracic legs and a limbless abdomen (Panganiban 2000). Abdominal *Dll* repression is caused by interactions with Ultrabithorax and Abdominal-A, Hox proteins of the Bithorax complex (BX-C) that bind the early *Dll cis*-regulatory element in combination with segment polarity gene products (En, Slp) and Pbc/TALE class proteins (Exd, Hth) (Gebelein et al. 2002; Gebelein et al. 2004; Merabet et al. 2003). The *Drosophila* early *Dll* regulatory network is summarised in Figure 1.2.1a overleaf, and detailed explanation is given in paragraphs which follow.

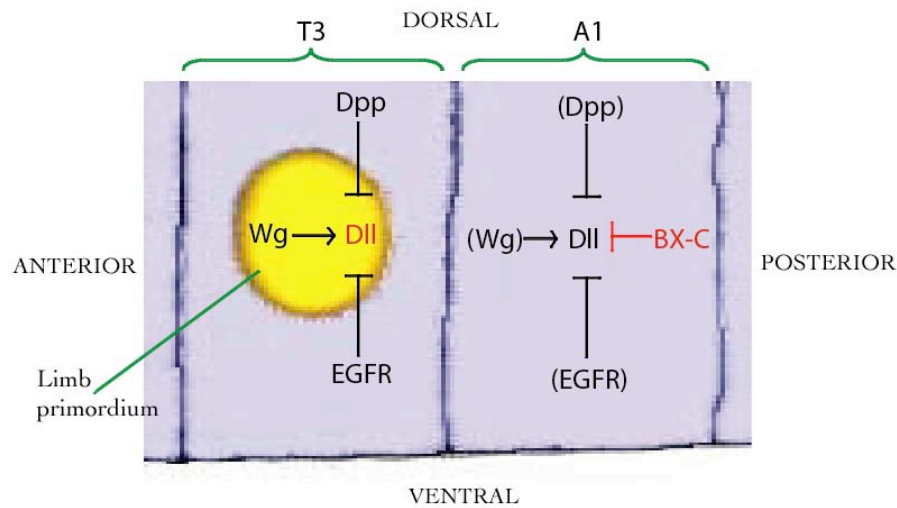


Figure 1.2.1a Summary of *Drosophila Distal-less* regulation in the thorax and abdomen, exemplified by segments T3 and A1. The output of positive and negative regulatory factors results in thoracic limb primordia, but limb suppression in abdominal segments. Wg: Wingless, Dpp: Decapentaplegic, EGFR: epidermal growth factor receptor, BX-C: Bithorax complex Hox proteins Ubx + AbdA (+ co-factors).

i) Antero-posterior segmental co-ordinates

In the cellularising *Drosophila* embryo, the segment polarity gene *engrailed* is activated in 14 single-cell wide transverse stripes (in subtle A-P and D-V progression), segments having been established by a well characterised hierarchy of genes ending in such segmentally reiterated pair-rule gene interactions (Akam 1987; DiNardo et al. 1985; Gilbert 2000; Karr et al. 1989). Activation of another segment polarity class gene *wingless*, in cells anterior to the *engrailed* row, creates an A-P parasegment boundary organising centre that is strictly maintained by an auto-regulatory positive feedback loop between Hedgehog (Hh), induced by En signalling, and Wg morphogens that confer a sharp anterior limit to *engrailed* stripes (see Figure 1.2.1b,i)(Gilbert 2000). Negative feedback control of *wg* also occurs, dependent on inducible inhibitors (e.g. *naked cuticle*, *nk2*) and cross-pathway antagonism (e.g. Hh-Zw3-GSK3 interaction)(Amit et al. 2002; Gerlitz and Basler 2002; Jia et al. 2002; Monnier et al. 2002; Rousset et al. 2001; Zeng et al. 2000; Zeng and Verheyen 2004). Once parasegments are established, *en* and *wg* are expressed independently of each other(Alexandre and Vincent 2003; Deutsch 2004; DiNardo et al. 1985; Ingham and Martinez Arias 1992; Jaynes and Fujioka 2004). Although both En and Wg signals are required for parasegmental organisation, it is transduction of the Wg signal that directly drives initial *Dll* expression from an 'early' *Dll*-304 enhancer located 5' of the gene, 12Kb upstream. After limb specification, *Dll* transcription is driven from a downstream

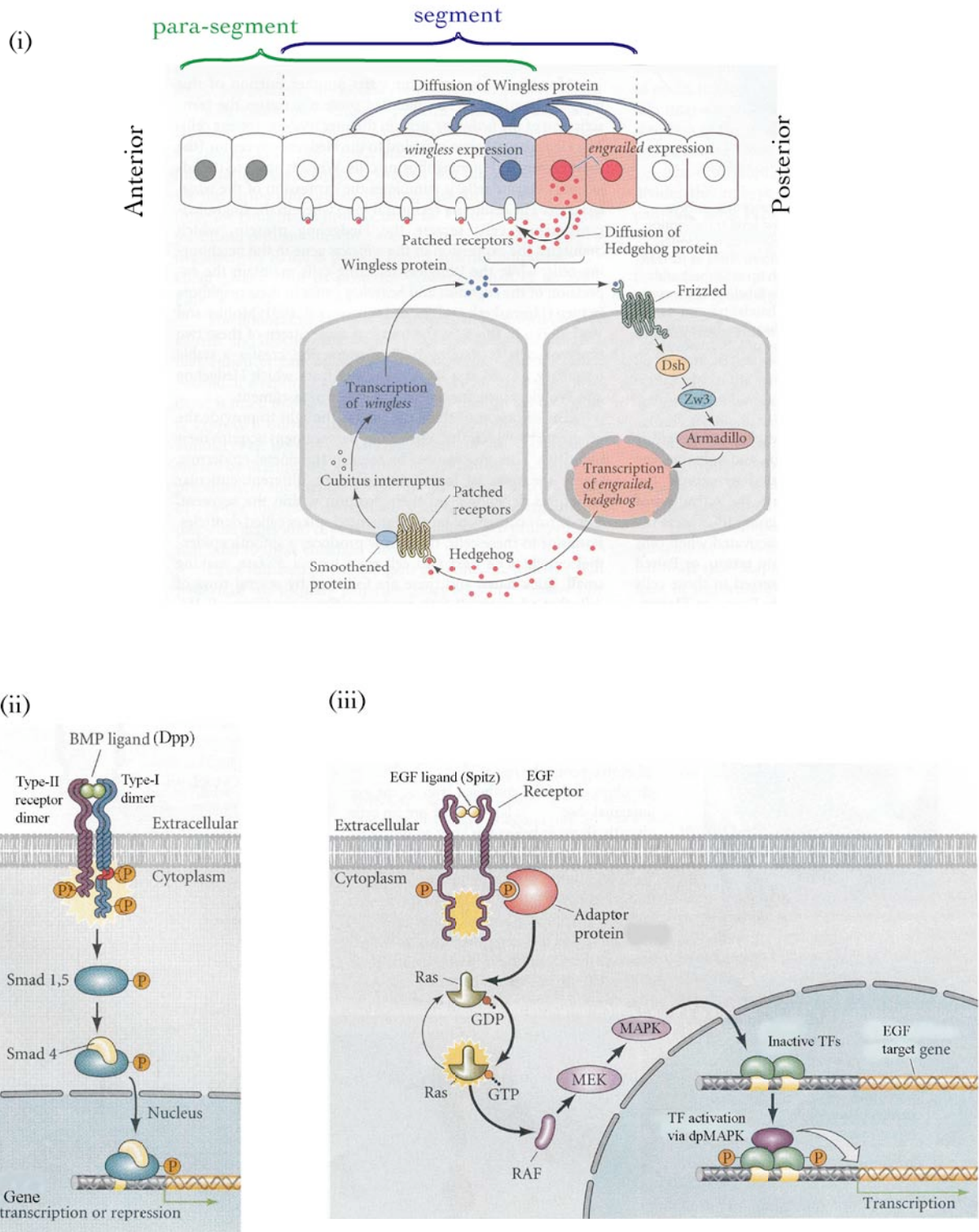


Figure 1.2.1b

Details of signaling pathways active at nodes within the *Drosophila* early *Dll* regulatory network: (i) En-Hh-Wg segment polarity gene network, (ii) Dpp/BMP (TGF-beta) signaling pathway; (iii) Canonical RTK-Ras-MAPK phosphorylation cascade induced by EGF signal transduction. Taken, with modified annotation, from Gilbert (2000) *Developmental Biology* Chapter 6.

'late' enhancer, independently of the initial Wg cues(Angelini and Kaufman 2005b; Cliffe et al. 2003; Kubota et al. 2003; Panganiban 2000).

ii) Dorsal-ventral co-ordinates

Transcription of the *Dll* gene is repressed dorsally by the Decapentaplegic (Dpp)-mediated Smad signalling pathway and repressed ventrally by Epidermal Growth Factor Receptor (EGFR)-mediated MAPK signalling(Gilbert 2000; Lewin 2002). Combined with segmental Wg that generates transverse stripes competent to express *Dll*, dorsal and ventral Dpp and MAPK repression limits final *Dll* transcription to paired, ventro-lateral clusters of approximately 20 cells(Cohen et al. 1993). These bilateral domains, genetically marked out by *Dll*, constitute limb primordia (Goto 1997, Klämbt 2001, Kubota, 2000).

Dorsal Decapentaplegic signals

The gene *decapentaplegic*, named after 15+ morphogenetic defects observed in original *dpp* mutants, encodes a TGF- β super-family signalling ligand that is characterised by a short signal sequence, precursor pro-domain and a mature, secreted peptide domain(Spencer et al. 1982). *Drosophila dpp* is homologous to, and functionally interchangeable with vertebrate *Bone Morphogenetic Protein 2 (BMP2)* and *BMP4* (Padgett et al. 1987; Padgett et al. 1993; Sampath et al. 1993). The mature Dpp ligand contains a 'cysteine knot' and one cysteine residue through which active dimers form by di-sulphide linkage. Activated Dpp functions as a morphogen in many developmental processes, *via* both short and long range trafficking mechanisms (Entchev et al. 2000; Tabata 2004). Dimeric Dpp ligands bind to Ser/Thr kinase domains of Type-II transmembrane receptors (e.g. Sax) that then phosphorylates a Type-I activin-like receptor kinase (e.g. Tkv), inducing cytoplasmic activation of a specific Smad isoform *via* a C' serine pair (Gilbert 2000; Lewin 2002; Persson et al. 1998; Suga et al. 1999). Activated Smad heterodimerises with the isoform Smad4, leading to nuclear transduction and transcription of target genes (for signalling pathway details, see Figure 1.2.1b).

Early zygotic *dpp* transcription is associated with establishing dorso-ventral polarity in the *Drosophila* blastoderm, and between 5-5.5hr AEL a dorsal gradient of Dpp signal emanating from the dorsal 40% of the embryo represses the early *Dll* enhancer, inhibiting *Dll* expression in cells exposed to high or intermediate levels of Dpp – i.e. cells located dorsally and dorso-laterally(Goto and Hayashi 1997).

Ventral EGFR activity

Drosophila EGFR (DER for short) is a receptor tyrosine kinase (RTK) whose activity is induced primarily by Spitz, the primary EGF ligand(Mayer and Nusslein-Volhard 1988; Rutledge et al. 1992; Shilo 2005). Spitz is chaperoned by Star in the ER, cleaved by Rhomboid and secreted extracellularly (Klämbt 2001; Tsruya 2002). It may diffuse up to 5 cell diameters from its source prior

to interacting with extracellular ligand-binding domains of EGFR, inducing receptor dimerisation (Shilo 2003; Shilo 2005). C' to each ligand binding domain are cysteine-rich domains containing diagnostic EGF repeats (CXCXXGF/YXGXXC) that create secondary structure *via* S-S bonds (Cooke et al. 1987; Gilbert 2000; Livneh et al. 1985). Short trans-membrane and basic juxta-membrane domains precede a long intracellular tyrosine kinase and auto-phosphorylation domain, including conserved ATP binding motifs (Greenwald 1985; Hsuan et al. 1989; Jorissen et al. 2003; Lepage et al. 1992). Auto-phosphorylation occurs upon dimerisation, inducing a cascade of cytoplasmic kinase activity ending in nuclear transduction of dual-phosphorylated mitogen-activated protein kinase (dpMAPK), which forms a complex with transcription factor(s) affecting target gene activity (for simplified illustration of the EGF signalling pathway, see Figure 1.2.1b) (Gilbert 2000). Secreted Spitz (sSpitz) acts as a ventral ectoderm determinant in the early *Drosophila* embryo, and after cellularisation and gastrulation EGFR activity is involved in limb specification and neural patterning, promoting differentiation and preventing apoptosis: From 2.5hrs AEL, Rhomboid is activated in response to a gradient of the maternal morphogen Dorsal, leading to autonomous Spitz secretion and graded EGFR activity in the ventro-medial ectoderm and mesectoderm (Ip et al. 1992). EGFR signalling interacts with medial Vnd and lateral Esg factors, specifying 3 rows of neuroblast cells along the dorso-ventral axis; EGFR is repressed medially, not essential laterally, but is critical in determining intermediate neuroblast identity (Shilo and Raz 1991; Skeath 1998; Udolph et al. 1998; Yagi et al. 1998). The product of another Spitz group gene, *single minded*/*sim*, is activated by Dorsal at the ventral midline, accompanied by adjacent proneural gene expression and responsible for ventral mesectoderm fate (Nambu et al. 1991). After cellularisation and gastrulation, from 3-4.5hr AEL, Sim induces Spitz secretion from a single row of cells at the ventral midline, diffusing from the midline to form a MAPK signalling gradient (Kim and Crews 1993). This signal targets *argos* (*aos*) transcription, and as Argos is an EGFR repressor, acting mainly by EGF ligand sequestration, this rapidly sets up a negative feedback loop, refining the MAPK gradient. Ventral EGFR and Dpp signalling antagonise one another along the D-V axis, mutually repressing each other's transcription. As the *Dll* gene is also a negative target of EGFR, dorsal and ventral domains of graded Dpp and EGFR-mediated repression result in restriction of *Dll* transcripts to a ventro-lateral focus within the antero-posterior Wg-dependent competence domain.⁵

⁵ Additionally at this stage, the MAPK gradient is modulated by Argos, guiding further neuroectoderm patterning as particular midline glial cells are selected for survival (Scholz et al. 1997; Stemerink 1997).

iii) Segment-specific Hox control

Hox gene expression along the antero-posterior axis in *Drosophila* and other arthropods is restricted to certain domains of the body, conferring specific appendage identity and morphology (Hughes and Kaufman 2002b). *Hox* genes of the Bithorax complex (BX-C), *Ultrabithorax* (*Ubx*) and *abdominal-A* (*abd-A*), are expressed in the *Drosophila* abdomen and their encoded proteins interact with the early *Dll* enhancer (via a number of co-factors), modifying the output of the generic, segmental antero-posterior and dorso-ventral limb-positioning cues described previously (Angelini and Kaufman 2005b; Gebelein et al. 2004).

Regulation and abdominal expression of BX-C genes

BX-C gene expression is controlled by *cis* and *trans* regulation involving both coding and, surprisingly, non-coding intergenic transcripts (the latter reviewed in Lemons & McGinnis, 2006) (Lemons and McGinnis 2006). Proposed regulatory mechanisms specific to *Ubx* and *abdA* transcripts include:

- (i) Recruitment of Trithorax group (TxG) chromatin remodelling proteins by intergenic non-coding RNA (e.g. ASH1 recruitment to maintain *Ubx* activation) (Schmitt et al. 2005).
- (ii) Transcription through non-coding polycomb group response elements (PREs), preventing PRE-mediated recruitment of repressive polycomb group (PcG) protein complexes and so preventing BX-C gene silencing (Schmitt et al. 2005).
- (iii) Loss and gain of micro-RNA target sites in BX-C regulatory elements (e.g. *mir-iab-4* miRNA binding represses *Ubx*) (Lemons and McGinnis 2006).
- (iv) Differential regulation of sense and anti-sense *Ubx* coding transcripts, mRNA expressed in complementary domains (e.g. in the *Drosophila* blastoderm) (Akam and Martinez Arias 1985).

Ubx is expressed in the cellular *Drosophila* blastoderm in a discrete parasegment (PS) 6 domain, from the posterior 3rd thoracic (pT3) – anterior 1st abdominal (aA1) segment. From gastrulation onwards, when the germband is forming and limb specification occurring, nuclear *Ubx* transcripts are detectable strongly from PS6 (pT3, dorsal) – PS12 (aA7) in ectoderm and mesoderm, and less strongly in ectoderm cells of PS5 (pT2) and PS13 (aA8). *Abd-A* expression is displaced one parasegment posterior of *Ubx*, covering a domain from PS7 (pA1) to PS14 (pA8) (Akam and Martinez Arias 1985).

BX-C activity in the Dll regulatory network

A minimal, 57bp *cis*-regulatory element ('*DllR*') responsible for early *Dll* repression in abdominal segments has been identified within the *Dll-504* element, and shown to contain binding sites for BX-C proteins, segmentation proteins and Pbc/TALE class co-factors, together forming a functional multi-

protein complex that affects repression (Figure 1.2.1c) (Gebelein et al. 2004; Vachon et al. 1992). Ubx and AbdA form trimeric complexes with Pbc class proteins Extradenticle (Exd) and Homothorax (Hth), mediated by a conserved Hox 'hexapeptide' motif (PRYPWM) N' to the homeodomain (HD) and stabilised by N' and C' residues either side (Merabet et al. 2003; Shanmugam et al. 1997). Hox/Pbc association increases Hox target DNA binding specificity, extending Ubx or AbdA target recognition motifs from TAAT (Hox) to TGATNNATNN (Hox/Exd/Hth).⁶ A degree of compartment-specific control over *Dll* regulation is achieved *via* combined integration of Ubx and AbdA inputs with segment polarity gene products on the minimal *DllR* repression element (Figure 1.2.1c). In functional assays, Ubx + Sloppy paired (Slp) are sufficient to repress *DllR* in anterior compartments, and AbdA + Engrailed (En) can repress *DllR* in posterior compartments, without an absolute requirement for Exd and Hth (Gebelein, 2004). Within the BX-C proteins themselves are non-HD regulatory modules that serve direct and indirect functions in affecting *Dll* repression in the abdomen and hence production of a limbless abdominal phenotype (Merabet et al. 2003; Pavlopoulos and Averof 2002). As mentioned previously, an N' hexapeptide motif mediates Hox/Exd interactions and Hox/Pbc complex stability on target regulatory elements, but in Ubx and AbdA other significant regulatory motifs are also present: Loss of ancestral serine/threonine residues C' to the Ubx HD is inferred to have released its repression capability from inhibition by ser/thr kinases, repression made more powerful by an acquired C' poly-alanine motif that enhances DNA binding (Galant and Carroll 2002; Ronshaugen et al. 2002). In the alternative splicing variant UbxIa, the 'linker' region connecting the YPWM motif and HD is involved in *Dll* repression; in AbdA, linker region residues interact with other intra-molecular domains to detect target genetic cues and mediate activation vs. repression responses (Gebelein et al. 2002; Merabet et al. 2003).

⁶ Proliferation of multiple Hox monomer binding sites (MMBSs) in *cis*-regulatory elements of target genes is also proposed to confer BX-C specificity (e.g. MMBSs for Ubx, leading to *Spalt* repression in T3), but this is an untested mechanism with respect to abdominal *Dll* repression (Galant 2002).

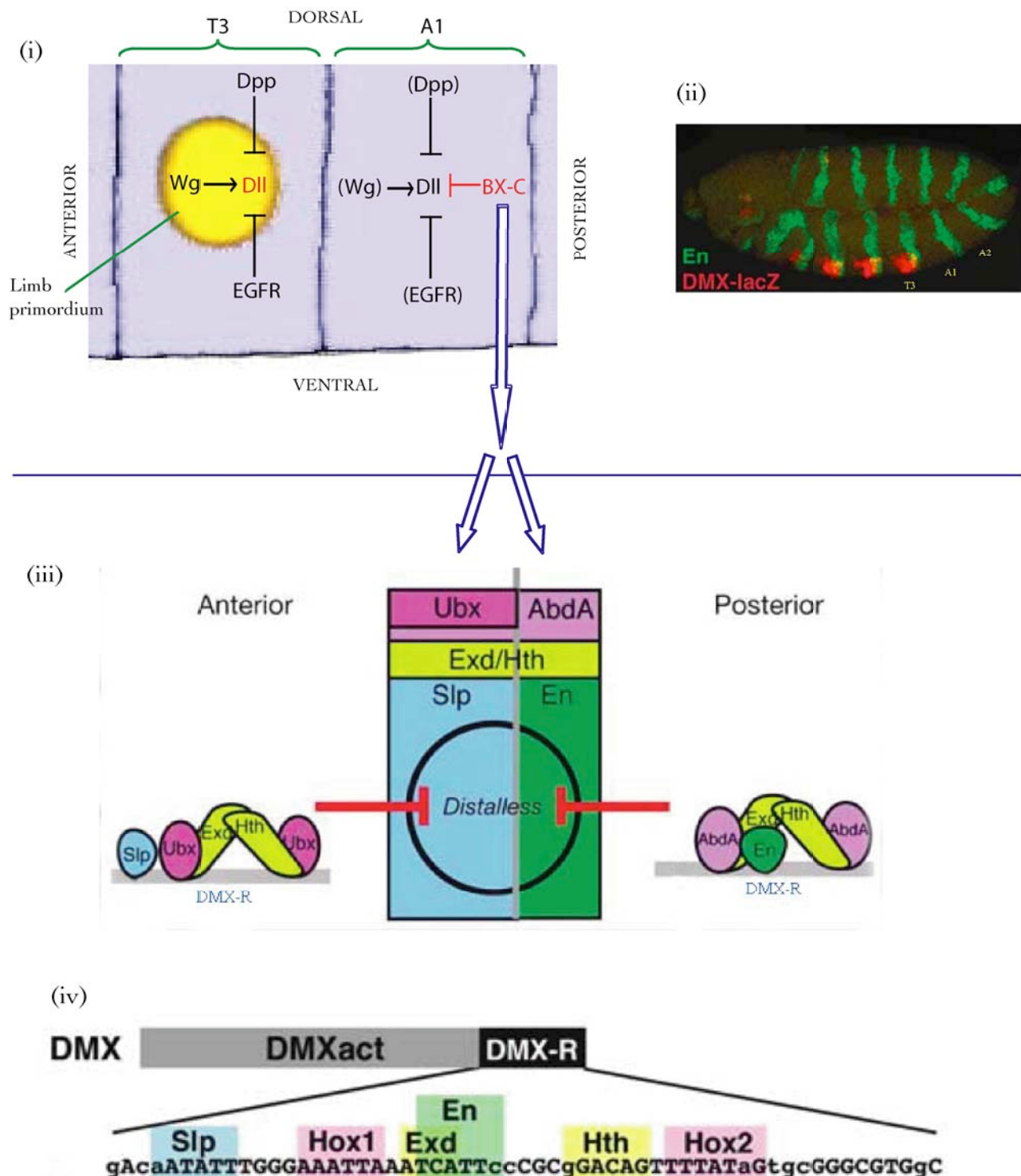


Figure 1.2.1c

- (i) Summary of *Drosophila* *Distal-less* regulation in thorax and abdomen segments, *Dll* active in the T3 limb.
(ii) Typical expression of *Dll* in T1-T3 limb primordia, driven by minimal enhancer 'DMX' (*Dll*-304); an 877bp element comprising large activator ('DMX-act') and small (57bp) repression ('DMX-R') domains.
(iii) Anterior vs. posterior compartment-specific interactions between Bithorax proteins Ubx and AbdA, segment polarity proteins Slp and En, and Exd/Hth DNA binding cofactors.
(iv) *Drosophila melanogaster* DMX-R nucleotide sequence, with binding sites for compartment localised transcription factors. DMX-R (57bp) is part of a larger 197bp general repressor element within *Dll*-304.

Note: (ii), (iii), (iv) from figures in Gebelein et al. (2004) Nature 431, 653-659.

1.2.2 The *Distal-less* regulatory network in divergent arthropods: moving beyond the *Drosophila* paradigm

i) Antero-posterior segmental signals

Patterns of *engrailed* gene expression, En-induced Hedgehog signalling and/or Hh-induced Wingless signalling have been studied in a wide range of insects and crustaceans, a few myriapods and chelicerates, and the onychophoran *Akanthokara kaputensis*.

During segmentation of the phylotypic arthropod germband, *en* orthologs and *wg/Wnt1* are consistently expressed just as in *Drosophila*, in single-cell wide ectodermal stripes demarcating anterior and posterior parasegment boundaries (c.f. Figure 1.2.1a)(Ingham and Martinez Arias 1992; Peel 2004; Scholtz 1997). Conserved activity of the En-Wg genetic circuit in defining and maintaining parasegmental boundaries raises the possibility of a broadly conserved mechanism for *Dll* activation during primordial limb allocation, which occurs once germband segmentation is stable(Damen 2002). However, the following factors and exceptions complicate this interpretation:

- i) recent loss-of-function phenotypes in three insect species,
- ii) presence of duplicated *en* and *Wnt* genes,
- iii) temporal dynamics of *En/Wg* and/or *Dll* expression in *Schistocerca americana* (orthopteran) and *Orchestia cavimana* (amphipod crustacean),
- iv) spatial dynamics of *wg/Wnt5* and *Dll* in *Myosidium colombiae* (malacostracan crustacean), and
- v) low taxonomic sampling outside the Pancrustacea.

Insect loss-of-function phenotypes RNA-interference⁷ (RNAi) experiments in *Oncopeltus fasciatus* (hemipteran) show that although *en* and *wg* are critical in segmentation, only *en* is required for proper limb development. This is surprising, given persistent expression of both genes parasegmentally and in a postero-ventral stripe along proximo-distal limb axes(Angelini and Kaufman 2005a). Similar results were obtained from RNAi-mediated gene knockouts in *Gryllus bimaculatus*, although *hh* RNAi (proxy for En activity) did not work(Miyawaki et al. 2004). By contrast, in the homo-metabolous coleopteran *Tribolium castaneum*, *wg* RNAi phenotypes revealed a need for *wg* in segmentation, limb specification and proximo-distal limb outgrowth, similar to conclusions made from analysis of

⁷ Powerful and specific gene silencing has been observed in both animals and plants, mediated by activation of an innate, immunity-related post-transcriptional RNA-induced silencing complex/RISC (Ahlquist 2002; Grishok 2001; Matzke 2005; Nishikura 2001; Xie 2003). Small, 21-25nt-long RNA molecules, dubbed micro-RNAs (miRNAs) or short interfering RNA (siRNAs), are synthesised mainly from double-stranded RNA (dsRNA) processing by the nuclease Dicer, and join a RISC complex which targets complementary mRNA for cleavage, degradation and hence silencing (Benfey 2003; Boutla 2001; Couzin, 2002; Elbashir 2001; Feinberg 2003; Henry 2004; Kennedy 2004; Kent 2004; Lipardi 2001; Scadden 2001). The result of this innate or artificially inducible (by introduction of synthetic dsRNA) mechanism is a cleanly gene-specific and hence highly informative knockout phenotype (Aljamali 2003; Baulcombe, D. 2002; Fire 1998; Hammond 2001; Henry 2003; Lam 2000).

Drosophila temperature-sensitive *wg* mutants (Cohen et al. 1993; Jockusch and Ober 2004). Given latest estimates of 12 *Wnt* subfamilies in an Urbilaterian ancestor, with cases of functional redundancy and co-option between extant family members, it may be that an unidentified non-*Wnt1* ortholog activates *Dll* and limb specification in *Gryllus* and *Oncopeltus*, but equally, *Wg/Wnt* may not have evolved a *Dll*-regulatory function until after divergence of hemi- and holo-metabolous insects (Kusserov et al. 2005).

Multiple *en* and *Wnt* family genes Tandem duplication to give paralogous Engrailed gene pairs has occurred in members of all four major arthropod sub-phyla, with redundant, sub-functionalised or differentially regulated gene activity linked to variously conserved roles in segmentation, neurogenesis and limb development. Duplication within the *Wnt* gene family and varying degrees of functional divergence between paralogs has been evident since before protostome and deuterostome divergence; for example, *Wnts* from cnidarians *Nematostella vectensis* and *Hydra vulgaris* point to at least 11 *Wnt* sub-families in basal metazoa, with retention and duplication to a maximum of 12 families in deuterostomes and loss of 5 or 6 families in protostome lineages (Hobmayer et al. 2000; Kusserov et al. 2005). Cnidarian *Wnt* expression, combined with data for more divergent metazoans (e.g. deuterostomes *Ciona savignyi*, *Lytechinus variegatus*), indicates that signalling induced by multiple *Wnt* family members was ancestrally associated with metazoan axis specification, followed later by widespread co-option to roles controlling cell fate and boundaries (Gordon and Nusse 2006; Imai 2003; Kusserov et al. 2005; Weitzel et al. 2004; Wikramanayake et al. 2003). Without certain knowledge about sequence and function of all *en* or *Wnt* paralogs in a given genome, there remains the possibility that the role of known orthologs is redundant, partially or completely fulfilled by an uncharacterised paralog.

Temporal *en/wg/Dll* dynamics In the short germband locust *Schistocerca gregaria*, *wg* transcription is observed significantly before *en* activation in anterior segments. This indicates that metameric patterning may be *en*-independent in the early germband, caused by divergent early *wg/en* regulation within anterior (gnathal) segments (Dearden and Akam 2001). Distinct segmental patterning mechanisms are postulated to have operated in anterior and posterior body regions either side of an ancestral arthropod boundary at/near PS4; locust *wg/en* transcription and translation dynamics may therefore have diverged in response to specific modification of early segmentation enhancers in anterior segments (Damen 2002; Peel et al. 2005). Complete temporal dissociation of *wg/en* circuitry from earliest *Dll* activation is exemplified by the amphipod crustacean *Orchestoidea cavimana* (Malacostraca), in which *Dll* is detected in primordial limb domains before *wg* or *en* gene activation (Hejnol and Scholtz 2004). This highlights a) the potential for significant species-specific regulatory evolution, as the isopod malacostracan *Porcellio scaber*, as well as a number of more

distantly related crustacea (e.g. decapods *Procambarus clarkii*, *Homarus americanus*, branchiopods *Artemia franciscana*, *Triops longicaudatus*) display ‘typical’ temporal *en/wg/Dll* dynamics and b) the benefit of broad intra- and inter-Order sampling (Abzhanov and Kaufman 2000a; Hejnol and Scholtz 2004; Patel et al. 1989).

Spatial en/wg/Dll dynamics Extreme divergence of *wg/Wnt1* function was found in another malacostracan crustacean, *Mysidium colombiae*, in which *wg* and *Wnt5* orthologs were expressed in some conserved regions (e.g. heart, muscle, eye) but absent from ventral neurectoderm and parasegmental domains inherited by limb primordia (Duman-Scheel et al. 2002). In the notostracan branchiopod *Triops*, a *wg* ortholog exhibits ‘incomplete’ parasegmental activity, expressed in ventral but not more lateral ectoderm stripes; similarly, complete parasegmental *wg*-like stripes are achieved in the spider *Cupiennius salei* only when *Cs-wg* is expressed complementary to its paralogue *Cs-Wnt5-1* (Damen 2002; Nulsen and Nagy 1999). These examples point to the potential for sub-functionalisation or co-option of non-*Wnt1* orthologs (or indeed non-*Wnt* family genes) into ancestral *Wnt1* roles.

Taxonomic sampling range Onychophoran Engrailed protein has been detected segmentally in the ectoderm, and also in somatic mesoderm and neural cells - indicative of ancestral expression in both segmental ectoderm and mesoderm derivatives, and possible signal induction between germ layers (Wedeen et al. 1997). Among non-Pancrustacean arthropods, mRNA transcripts for chelicerate and myriapod orthologs of *en*, *wg*, *hb* and the activator of Hh signalling *cubitus interruptus* (*Ci*) indicate critical conservation of a parasegmental En-Wg genetic regulatory circuit in germband segmentation processes. Expression data are available for *Cupiennius salei* (spider), *Euscorpius flavicaudis* (scorpion), *Lithobius atkinsoni*, *Strigamia maritima* (chilopods), *Ethmostigmus rubripes* and *Glomeris marginata* (diplopods), although no functional evidence exists so far (Chipman et al. 2004a; Chipman et al. 2004b; Damen 2002; Hughes and Kaufman 2002a; Kettle et al. 2003; Prpic 2004a; Prpic 2004b; Prpic et al. 2003; Simonnet 2005; Simonnet et al. 2004; Whittington et al. 1991). Given conservation of *wg* expression but not function during limb specification and outgrowth in hemimetabolous insects (e.g. *Gryllus* and *Oncopeltus*), loss-of-function data for myriapods and chelicerates are very desirable to confirm or refute effects of AP En/Wg signals on early *Dll* enhancers and limb specification.

I propose to obtain gene sequence and expression data for *Tetranychus urticae en* and *wg/Wnt* orthologs, in order to extend available information on conservation or sub-functionalisation of chelicerate parasegment genes (c.f. section 1.5). Unlike *Euscorpius flavicaudis* and early stage *Cupiennius salei* embryos, the spider mite may also represent a species amenable to generation of RNAi phenotypes (c.f. section 1.4), permitting true functional analysis.

ii) Dorso-ventral patterning signals

∂pp transcription and signal activity has been examined in insects, the myriapod *Glomeris marginata* and a number of chelicerate species, but in no arthropods other than *Drosophila melanogaster* has EGFR activation been described or analysed. Although aspects of Dpp signalling relating to dorsal blastoderm, extra-embryonic and proximo-distal limb patterning may be conserved among the arthropods, it seems that outside *Drosophila* there is no strictly conserved requirement for *∂pp* to directly restrict *Dll* transcription during limb allocation.

Dpp in early axial patterning Within holometabolous, endopterygote insects *∂pp* is first expressed in the dorsal blastoderm (e.g. *Drosophila melanogaster*, *Athalia roseae*, *Tribolium castaneum*), associated with specifying extra-embryonic membranes and the dorsal embryonic axis (Sanchez-Salazar et al. 1996; St Johnston and Gelbart 1987; Yamamoto et al. 2004). Within hemimetabolous neopterans, *∂pp* is associated with posterior germband invagination (e.g. *Oncopeltus fasciatus*) and expressed in dorsal extra-embryonic and posterior germband territories (e.g. *Schistocerca americana*, *S. gregaria*), indicative of putative ancestral Dpp activity in early dorsal and axial fate specification (Dearden and Akam 2001; Jockusch et al. 2000; Jockusch et al. 2004; Niwa et al. 2000). In the arachnomorph *Achaearanea tepidariorum* and less derived haplogyne spider *Pholcus phalangoides*, *∂pp* activity has been connected with early axial patterning of an initial, radially symmetrical germdisc (Akiyama-Oda and Oda 2003; Akiyama-Oda and Oda 2006). Dpp signals from migrating cumulus mesenchyme cells specify dorsal identity in overlying ectoderm, antagonised by complementary ventral *Sog* expression in a primordial ventral domain; Dpp also induces extra-embryonic cell identity at the posterior pole, indirectly affecting anterior pole identity at the most opposite point (Akiyama-Oda and Oda 2006). This adds to findings in other chelicerates (Xiphosurans *Limulus* and *Tachypleus*; spider *Agelena labyrinthica*), where secondary dorso-ventral axes develop after grafting ectopic ‘primitive cumulus’ cells either within or between species, due to effectively transplanting axial ‘organiser’ properties (Akiyama-Oda and Oda 2003; Holm 1952; Itow et al. 1991). Hence, a role for *∂pp* in extra-embryonic and/or axial patterning is inferred as ancestral not only for hexapods, but for all arthropods.

Dpp in germband segmentation and limb specification During early germband stages of both the hemipteran *Oncopeltus fasciatus* and chelicerate *Cupiennius salei*, *∂pp* mRNA localises to transverse segmental stripes, each anterior to a parasegment boundary: stripe position overlaps spatially with parasegmental *Wg* activity, and in *Cupiennius* stripes fade at the ventral midline (Angelini and Kaufman 2005a; Prpic et al. 2003). These data appear to correspond to a time-point when limb primordia would be being specified, and if so, might suggest an ancestral mechanism of *Dll* regulation that has divergent genetic requirements and dynamics relative to *Drosophila*. By contrast, in the germbands of several lower and higher insects (apterygote *Thermobia domestica*, orthopterans

Schistocerca gregaria, *S. americana*⁸, *Gryllus bimaculatus*, coleopteran *Tribolium castaneum*), longitudinal *∂pp* stripes mark dorsal-most ectoderm, similar to the scenario in *Drosophila* (Giorgianni and Patel 2004; Goto and Hayashi 1997; Jockusch et al. 2000; Niwa et al. 2000). However, none of these longitudinal stripes become modulated as in *Drosophila*; no novel antero-dorsal clusters or antero-dorsal bands are reported, that in *Drosophila* would restrict *Dll*-positive cells. Similarly, in the myriapod *Glomeris*, *∂pp* in the germband is first expressed in longitudinal stripes of dorsal neurectoderm (later also transiently at the ventral midline), but any modulation hypothetically connected to limb specification is not evident (Prpic 2004a). Clearly, it would be beneficial to view *∂pp* expression in chelicerate, myriapod and crustacean taxa, in the temporal window specific to limb primordium allocation, so that the functional significance of Dpp signals, segmental or axial, can be more fairly assessed.

Dpp in proximo-distal and dorso-ventral limb axes In chelicerates, myriapods and insects whose limbs grow outward from axial buds (rather than imaginal discs as in *Drosophila*), Dpp signalling in appendages is first active in distal limb bud domains. Early distal domains are restricted to the dorsal side in some cases (e.g. *Tribolium castaneum* antenna, *Glomeris marginata* legs), but otherwise appear to include both dorsal and ventral tip cells (Giorgianni and Patel 2004; Prpic 2004a). Later *∂pp* expression is dynamically modulated along the proximo-distal axis, common features including dorsal stripes, patches and both inter- and intra-segmental circumferential rings (Prpic et al. 2003). A homologous mechanism of arthropod proximo-distal limb patterning was proposed by Prpic (2004) on the basis of conserved Wg and Dpp morphogen concentration gradient topologies between the two-dimensional *Drosophila* imaginal disc and three-dimensional buds of other arthropods (Prpic 2004a). However, a recent functional study in the coleopteran *Tribolium castaneum* confounded this ‘topology hypothesis’, as the *Tribolium ∂pp* RNAi phenotype has disrupted dorsal extra-embryonic ectoderm but no defects in dorso-ventral or proximo-distal appendage development, indicating that *∂pp* is either redundant or not used in limb specification and proximo-distal patterning (Jockusch and Ober 2004; Jockusch et al. 2004). This result was surprising given persistent Dpp signalling activity in *Tribolium* limbs, and as well as implying *∂pp* co-option to proximo-distal limb patterning after coleopteran-dipteran divergence it draws attention to the need for functional analyses outside the holometabola⁹ to test the *in vivo* significance of *∂pp* transcription dynamics.

⁸ In all cases except *Schistocerca americana*, published *∂pp* gene expression data correspond to phases just beyond limb specification, when proximo-distal limb development has commenced.

⁹ The loss-of-function phenotype for *Oncopeltus fasciatus ∂pp* is uninformative regarding appendage development, as the germband fails to invaginate due to loss of posterior germband *∂pp* function. Progress to limb specification and development stage was similarly prevented in *Achaearanea tepidariorum ∂pp* RNAi embryos, where failed embryonic dorso-ventral axis specification severely compromises germband patterning.

iii) Segment-specific *Hox* control

Hexapoda (Protura + Collembola + Diplura + Insecta)

In insects, abdominal appendages are suppressed through *Distal-less* repression by the Bithorax proteins Abdominal-A and - in holometabolous lineages – Ultrabithorax, resulting in the typical hexapod bodyplan with limbed thorax and limbless abdomen (Figure 1.2.2a)(Angelini and Kaufman 2005b; Panganiban and Rubenstein 2002). In the basal hexapod group Collembola, Ubx and AbdA ('UbdA') are co-expressed with Dll in the abdomen, indicating lack of repressive interaction. *Dll*-expressing ventral appendages of different forms develop on abdominal segments A1 (ventral tube/collophore), A3 (retinaculum) and A4 (furca) in Collembola, but A2, A5 and more posterior segments are limbless: BX-C gene-independent mechanisms must exist to specify appendage identity or mediate limb repression¹⁰, but as yet such mechanisms are unknown (Gullan and Cranston 2000; Palopoli and Patel 1998). After divergence from Collembola (Figure 1.2.2a), AbdA evidently acquired the ability to repress *Dll*, either by altered Hox protein properties or addition of AbdA repression-specific binding sites in *Dll cis*-regulatory sequences, repressing limb development in all abdominal segments where AbdA is present (caudal of pA1). Similarly, Ubx acquired ability to modify A1 appendage morphology, forming the reduced pleuropod appendage characteristic of apterygotes such as *Thermobia domestica* (Thysanura) and hemimetabolous pterygote taxa such as *Tribolium castaneum* (Coleoptera) and *Oncopeltus fasciatus* (Hemiptera): in these lineages Ubx does not repress *Dll* appendage development, but affects target genes downstream of the appendage allocation gene network (Angelini et al. 2005; Galant and Carroll 2002; Lewis et al. 2000). After divergence of hemi- and holometabolous insects (~325Ma), Ubx evolved to completely repress limb development, partly explained by loss of C' domain S/T kinase sites (hitherto implicated in inhibiting repression), to be replaced by a poly-A track that enhances DNA binding and mediates powerful target gene repression (Figure 1.2.2a)(Vervoort 2002; Walldorf et al. 2000). Certain holometabolous insects possess larval abdominal prolegs, formed by either localised BX-C gene repression that allows *Dll* activation in abdominal limb primordia (e.g. lepidopterans *Manduca*, *Bombyx*, *Precis* where Ubx-AbdA 'holes' appear in segments A3-A6 due to secondary down-regulation), or through deploying a BX-C/*Dll*-independent genetic program for limb development (e.g. symphytid hymenopterans *Neodiprion* and *Diprion* where prolegs represent proximal limb segment(s) only) (Palopoli and Patel 1998; Suzuki and Palopoli 2001; Ueno et al. 1992; Zheng et al. 1999).

¹⁰ This conclusion is conditional upon the assumption that BX-C factors are not subject to significant temporal regulation at the point of limb specification, as they are in *Drosophila*. Spatial modulation of *Drosophila* Ubx in thoracic segments is critical to permitting development of T2 and T3 appendages, and *Ubx/abdA* expression is independent of *Dll* during later limb growth (Castelli-Gair et al.1994)

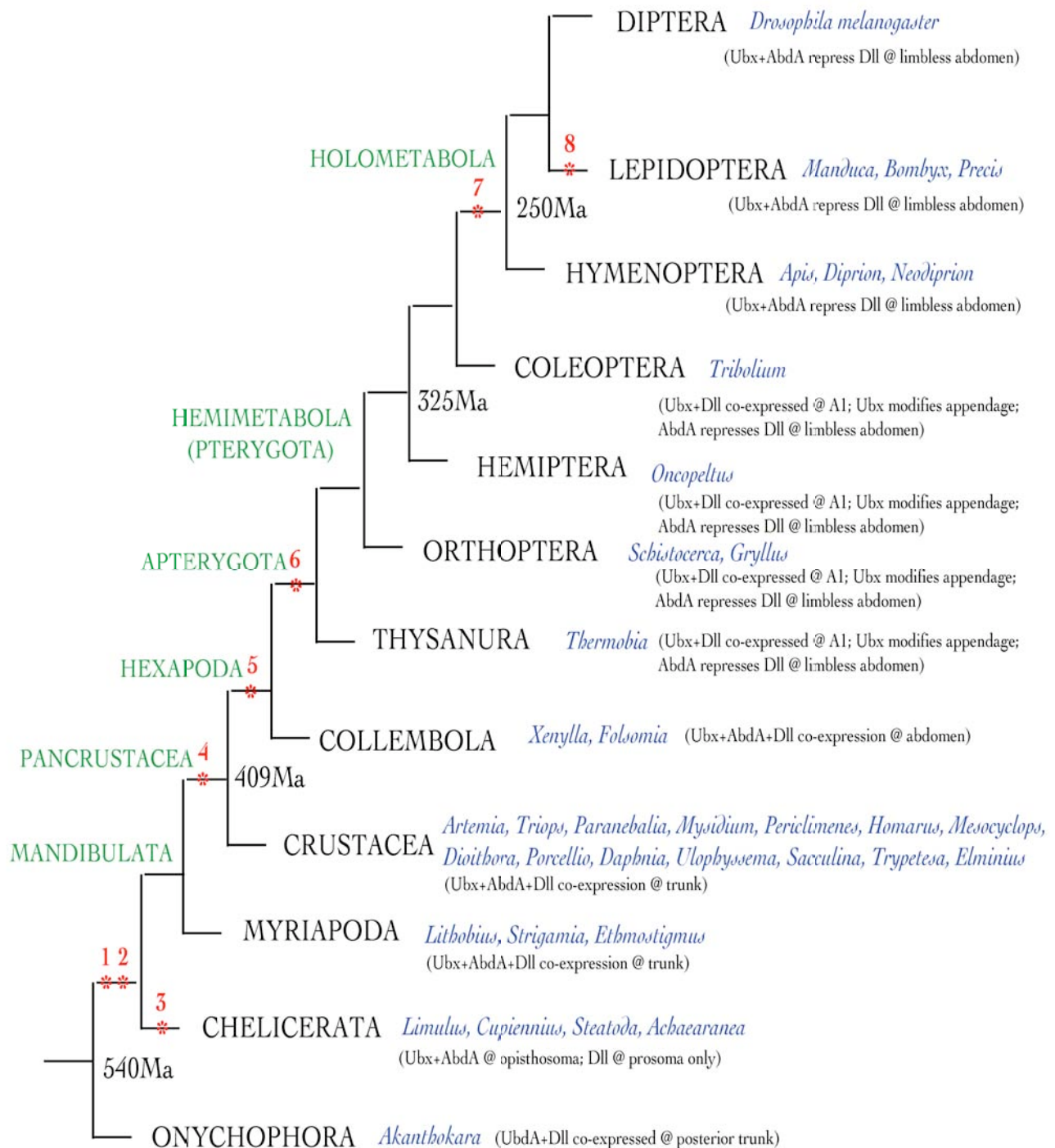


Figure 1.2.2a Molecular evolutionary features regarding the evolution of repression between Ubx, AbdA and Distal-less.

- 1: Anterior expansion of *UbdA* expression domain into trunk [*Ubx* + *abdA* evolution *in cis*].
- 2: Ubx C' QAQAAKAAA motif; phosphorylation (phphn) sites, domainly GSK3-kinase type [*Ubx* evolution *in trans*].
- 3: Hypothesis = *Dll* repression in the chelicerate opisthosoma. Q = UbdA dependent or independent mechanism?
- 4: Ubx C' domain inhibits ability of Ubx to repress *Dll*; C' addition of CKII-type phphn sites [*Ubx* evolution *in trans*].
- 5: Loss of Ubx C' phphn sites [*Ubx* evolution *in trans*]. Ubx & AbdA unable to repress *Dll* in collembolan abdomen.
- 6: Ubx modifies A1 limb to pleuropod. Ubx C' QAQAQKAAA+poly-A motif; Ubx C' loss phphn sites [*Ubx* evolution *in trans*]. AbdA represses *Dll* in abdomen. Either (i) AbdA acquires repression ability [evolution *in trans*], or (ii) Repressive AbdA binding sites appear in early *Dll cis*-regulatory element [evolution *in cis*].
- 7: Ubx represses *Dll* in A1 segment = derived Ubx state. Evolution *in trans* (e.g. C' motif) or *in cis* (e.g. *Dll* binding sites for repressive Ubx)?
- 8: Secondly derived changes *in cis* affecting *Ubx* + *abdA* expression: localised *BX-C* repression --> *Dll* de-repression --> abdominal prolegs.

Crustacea + Myriapoda

In non-hexapod mandibulates (i.e. Crustacea + Myriapoda), UbdA and Dll proteins are consistently co-expressed during specification of diverse limb types of corresponding trunk segments. Functional studies in the branchiopod crustacean *Artemia franciscana* have highlighted the importance of Ubx C' protein sequence evolution, in that *Artemia* Ubx contains multiple C' S/T kinase phosphorylation sites, implicated in inhibition of a latent repression domain elsewhere N' of the Ubx HD (Ronshaugen et al. 2002). However, given variation in the degree with which inferred ancestral arthropod C' Ubx residues (characterised by a QAQAAXAA motif) are replaced by S/T kinase phosphorylation sites in both crustacean and myriapod Ubx sequences (see Figure 1.2.2b for Ubx C' multiple alignment), the model may be conditional on a lower threshold level of S/T kinase activity than in the rather exceptional, as it turns out, *Artemia franciscana* C' domain. Alternatively, altered *Dll cis*-regulatory BX-C response elements and/or as yet unknown protein changes in both Ubx and AbdA protein modules may be required to explain the lack of *Dll* repression by UbdA in distinct Crustacean and Myriapod lineages (Averof 2002; Galant et al. 2002; Hsia and McGinnis 2003).

Chelicerata

Within the Chelicerata, UbdA proteins are consistently expressed in opisthosomal segments, whereas early *Dll* and ventral limbs are notably absent throughout this posterior body region, being restricted to anterior, prosomal segments only (Abzhanov et al. 1999; Cornec and Gilles 2006; Popadic and Nagy 2001). Exclusive UbdA and Dll domains are reminiscent of the situation in higher insects where BX-C factors directly repress the *Dll-504* regulatory element, begging the question whether chelicerates have independently evolved the same mechanism for BX-C-mediated *Dll* repression as insects, or whether novel regulatory mechanisms underlie the convergent phenotype of limbless posterior tagma in both groups. As far as concerns Ubx function, chelicerate Ubx C' domains are generally rich in S/T phosphorylation sites and although they conserve the arthropod QAQAAXAA motif to varying degrees, they lack the poly-A track characteristic of insects, thus indicating lack of analogous evolution at the level of Ubx C-termini (c.f. Figure 1.2.2b for Ubx C' multiple alignment). If chelicerate BX-C genes do in fact repress *Dll* in the opisthosoma - which requires loss-of-function evidence to prove - target gene regulation dependent on i) as yet unknown Ubx and/or AbdA protein modules or ii) novel *Dll cis*-regulatory control elements will have to be explored.

Onychophora

As a close outgroup to the Arthropoda, Onychophora provide a means to assess ancestral character states and inform the direction of molecular (and morphological) evolution. In the homonomous,

	1	20	40	60	
Dm-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Dv-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Tc-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Jc-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Am-Ubx	MAHSLCLTERQ	I	KI	WFQ	NR
Fc-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Af-Ubx	MAHSLCLTERQ	I	KI	WFQ	NR
Sc-Ubx	MAHQLCLTERQ	I	KI	WFQ	NR
Ps-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Tc2-Ubx	MAHQLCLTERQ	I	KI	WFQ	NR
Er-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Ag-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Gm-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
La-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Sm-Ubx	MAHALCLTERQ	I	KI	WFQ	NR
Cs-Ubx1	MAHALCLTERQ	I	KI	WFQ	NR
Cs-Ubx2	MAHSLCLTERQ	I	KI	WFQ	NR
Ak-Ubx	MAHALCLTERQ	I	KI	WFQ	NR

Figure 1.2.2b Multiple alignment of arthropod/onychophoran Ubx C terminal regions, including partial homeodomain (GREY), Ubx signature peptide (PURPLE) and final C terminal residues. BLUE: shared identity to *Drosophila/Junonia* QAQAK+10x/12x alanine motif. Red font: all sites of putative ser/thr kinase mediated phosphorylation. YELLOW: MAP kinase phosphorylation consensus site (thr-pro/ser-pro). GREEN: casein kinase II (CKII) consensus phosphorylation site (ser/thr-X-X-asp/glu). Boxed: GSK-III consensus phosphorylation sites, overlapping where multiple in succession (ser/thr-X-X-ser/thr).

Species: Dm = *Drosophila melanogaster*; Dv = *Drosophila virilis*; Tc = *Tribolium castaneum*; Jc = *Junonia coenia*; Am = *Apis mellifera*; Fc = *Folsomia candida*; Af = *Artemia franciscana*; Sc = *Sacculina carcini*; Ps = *Porcellio scaber*; Tc2 = *Trigriopus californicus*; Er = *Ethmostigmus rubripes*; Ag = *Archisporotreptus gigas*; Gm = *Glomeris marginata*; La = *Lithobius atkinsoni*; Sm = *Strigamia maritima*; Cs = *Cupiennius salei*; Ak = *Akanthokara kaputensis*.

lobopod-bearing trunk of *Akanthokara kaputensis*, Ubx and AbdA protein is restricted to the posterior-most lobopod segment only (Grenier et al. 1997). *Dll* is expressed in all trunk limbs, including lobopods in the Ubx domain, indicating lack of repressive interaction between BX-C and *Dll* before divergence from true arthropods ~540Ma (and given morphological homonymy of all lobopods, probably also lack of BX-C regulation of gene batteries affecting limb differentiation) (Grenier et al. 1997). Specifically regarding Ubx and the model of C' domain evolution, the QAQAAXXAA and phosphorylation motifs characteristic of basal arthropods are absent from *Akanthokara kaputensis* Ubx, indicating that these motifs are synapomorphies for Arthropoda, present since divergence of the common chelicerate + mandibulate ancestor from the Onychophora (c.f Figure 1.2.2a).

1.2.3 An extra player? Regulation of limb specification by *Sp* genes

Sp proteins¹¹ were first isolated in the late 1980s, but only in recent years has it emerged that certain *Sp* family genes play roles in appendage development that may be conserved in the Arthropoda, and even throughout the Bilateria (Trieckel et al. 2003). Before considering the evidence for *Sp* gene function in disparate animals' appendages, I shall briefly outline the diagnostic characteristics and evolution of the *Sp* gene family, as an appreciation of paralogy groups is important when considering accounts of functional diversification vs. redundancy among *Sp* duplicates:

Sp factor properties

Sp genes encode proteins related to Krüppel-like factors; both share a C terminal DNA binding domain comprising 3 conserved Cys2His2 Zn fingers that recognise GC/GT boxes within many transcription factor (TF) promoters (GC box = GGGGCGGGG; GT box = GGTGTGGGG) (Athanikar et al. 1997; Bouwman and Philipsen 2002). *Sp* genes are uniquely distinguished by i) an N terminal *Sp* box (SPLALLAATCS[R/K]I), implicated in regulating *Sp* factor proteolysis or repressor-mediated trans-activation, and ii) a 10 amino acid Buttonhead (Btd) box N' terminal to the zinc finger (ZnF) domain, synergising or involved directly in transactivation (Athanikar et al. 1997; Bouwman and Philipsen 2002; Kadonaga et al. 1987). Different *Sp* factors diverge in sequence outside the ZnF domain, conferring distinct properties and transcriptional regulatory functions, although partial redundancy in paralogous *Sp* function does occur in some contexts (Bell et al. 2003; Bouwman and Philipsen 2002).

¹¹ Nomenclature: *Sp* genes are known as 'specificity protein' genes, but the name originally derives from the sephacryl and phosphocellulose columns used for their initial purification (Bouwman 2002; Beermann 2003; Kadonaga et al. 1987).

Sp gene evolution

The current model of *Sp* gene evolution (Figure 1.2.3) infers tandem duplication of a single ancestral gene to form an ancestral Urbilaterian *Sp8*-type + *Sp1/5/4*-type gene pair (Kawakami et al. 2004; Wimmer 2005). These two genes are recognisable as distinct orthologs in many arthropod lineages, including members of all major sub-phyla except – until now – the chelicerates (c.f. Figure 1.2.3). The *Sp* gene complement of Dipterans is modified, as an *Sp8*-type gene duplication has produced the paralog *buttonhead/btd* (e.g. in *Drosophila melanogaster* and *Clogmia bipunctata*), and the representative *Sp1/5/4*-type gene seems to have been lost in certain lineages (e.g. *Clogmia*) (Wimmer et al. 1996; Wimmer et al. 1995). Orthologs of *Drosophila Sp1/5/4* and *btd* have not been identified in *Tribolium castaneum*, indicating either that the genes await discovery, or that the *Sp8-btd* duplication occurred at the base of the Diptera rather than the base of the Holometabola or Hexapoda, and loss of *Sp1/5/4* orthologs has occurred independently several times (Beermann et al. 2004; Janssen 2005; Wimmer 2005). In the vertebrates, whole genome duplications affecting chromosomes carrying the original two progenitor genes produced *Sp* genes arranged in pairs as *Sp1+Sp7/Osterix*, *Sp2+Sp6*, *Sp3+Sp9* and *Sp4+Sp8*, adjacent to *Hox* clusters C, B, D and A respectively. *Sp5* is associated with *Sp5* in *Homo*, is unlinked to other *Sp* genes in some other mammals, and has a modular genomic structure affiliating it with the *Sp6-9* subfamily (see Figure 1.2.3) (Bouwman and Philipsen 2002; Kawakami et al. 2004; Thorpe et al. 2005; Weidinger et al. 2005).

Sp gene function in appendage development

In *Tribolium castaneum*, *Sp8* expression is uniform in early limb buds, then restricted to ring-like expression domains along the P/D axis that activate Notch signalling, controlling cell proliferation and allometric growth of the limbs. *Sp8* operates upstream of *Dll*, its localised absence in *Tribolium* preventing formation of *Dll*-dependent structures (Beermann et al. 2004). Data for the spider *Cupiennius* also points to a role for Notch in limb development that could be linked to *Sp* gene activity, and the presence of stunted limbs when *Dll* function is removed by RNAi points to genes other than *Dll* being at least partly responsible for, or capable of inducing, some form of limb development (Schoppmeier and Damen 2001; Stollewerk et al. 2003a). Recent work in *Drosophila* has confirmed partially redundant roles for the *Sp8/btd* gene pair in ventral appendage formation: *Sp8/btd* responds to Wg, Dpp and BX-C factors, activating *Dll* and *btd/exd* to define distal and proximal ventral imaginal disc primordia (Estella et al. 2003). Remarkably, *Sp8* and *btd* are also capable of over-riding dorsal imaginal disc signals when expressed ectopically, converting dorsal to ventral appendage identity by redeploying the genetic network for ventral limb specification and outgrowth (Estella et al. 2003; Morata and Sánchez-Herrero 1999). In vertebrates, most reports of *Sp* gene roles relate to maintaining signals during earliest limb outgrowth (e.g. FGF, Wnt and BMP at the apical ectodermal

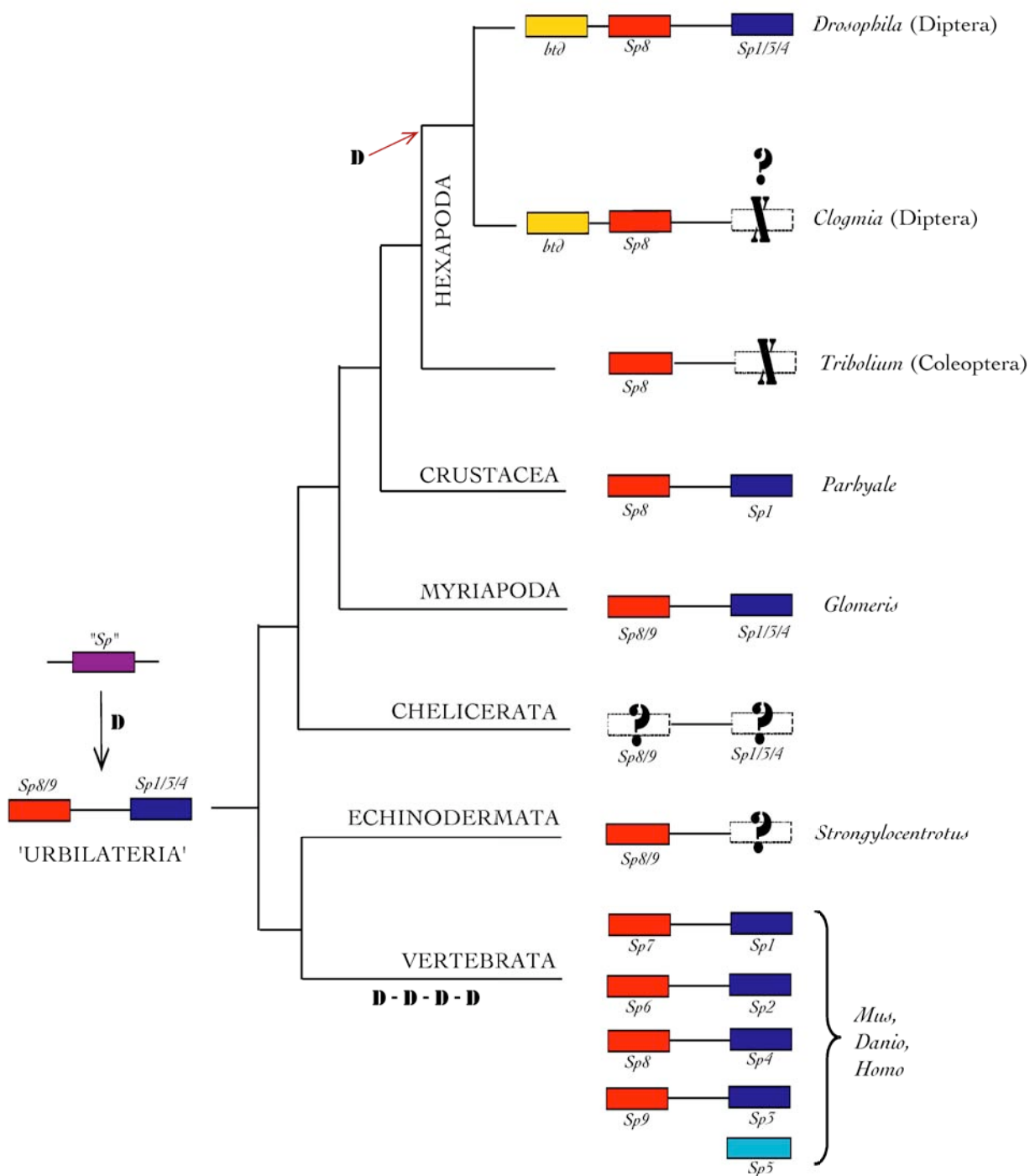


Figure 1.2.3 Model of *Sp* gene family evolution, based on available protostome (arthropod) and deuterostome data. Tandem duplication (**D**) of an original "*Sp*-like" gene produces one *Sp8/9*-type + one *Sp1/5/4*-type in a hypothetical urbilaterian ancestor. The gene pair are retained in arthropod lineages. Further tandem duplication of *Sp8* in Diptera given the gene *buttonhead*, and some Dipteran lineages appear to display instances of *Sp1/5/4* gene loss (**X**). It is not yet clear (?) if the *Sp1/5/4* gene lost in *Tribolium* has been lost independently in *Strongylocentrotus*. Four whole genome duplications occurred in the vertebrate lineage, giving rise to the two large sub-groups illustrated simply in the tree diagram.

Figure based on: Wimmer et al. (1995, 1996, pers. com.); Estella et al. (2003); Beermann et al. (2003); Janssen (pers.com.); Kawakami et al. (2004); Panopoulou & Poutska (2005); *Tribolium* genome project.

ridge) and anterior neurogenesis (Bell et al. 2003; Triechel et al. 2003; Weidinger et al. 2005). Knocking out *mBtd* (*Sp8*) function in mice causes limb truncation in a manner similar to *Tribolium*, the limbs often retaining structural integrity but displaying stunted growth (Triechel et al. 2003). As evidence in vertebrates supports a role for *Sp8/btd* in maintaining both P/D and D/V limb patterning, combined with arthropod data this suggests a possible ancestral role in metazoan appendage differentiation, perhaps *via* interactions with Notch-type signalling cues. Of note to this project, *Drosophila Sp8/btd* clearly also retains potential to regulate specification of ventral appendage fate, acting as an intermediate between axial co-ordinate signals and *Dll* transcriptional activation. Hence, one of my aims (section 1.5) became to test for conserved expression of any *Sp8*-like ortholog I might identify in *Tetranychus*, adding critical basal arthropod data to the emerging picture of *Sp* functions in arthropod limb specification and proximo-distal development.

1.3 Choice of an organism: *Tetranychus urticae*

1.3.1 Arthropod phylogeny – pertinence of the Chelicerata

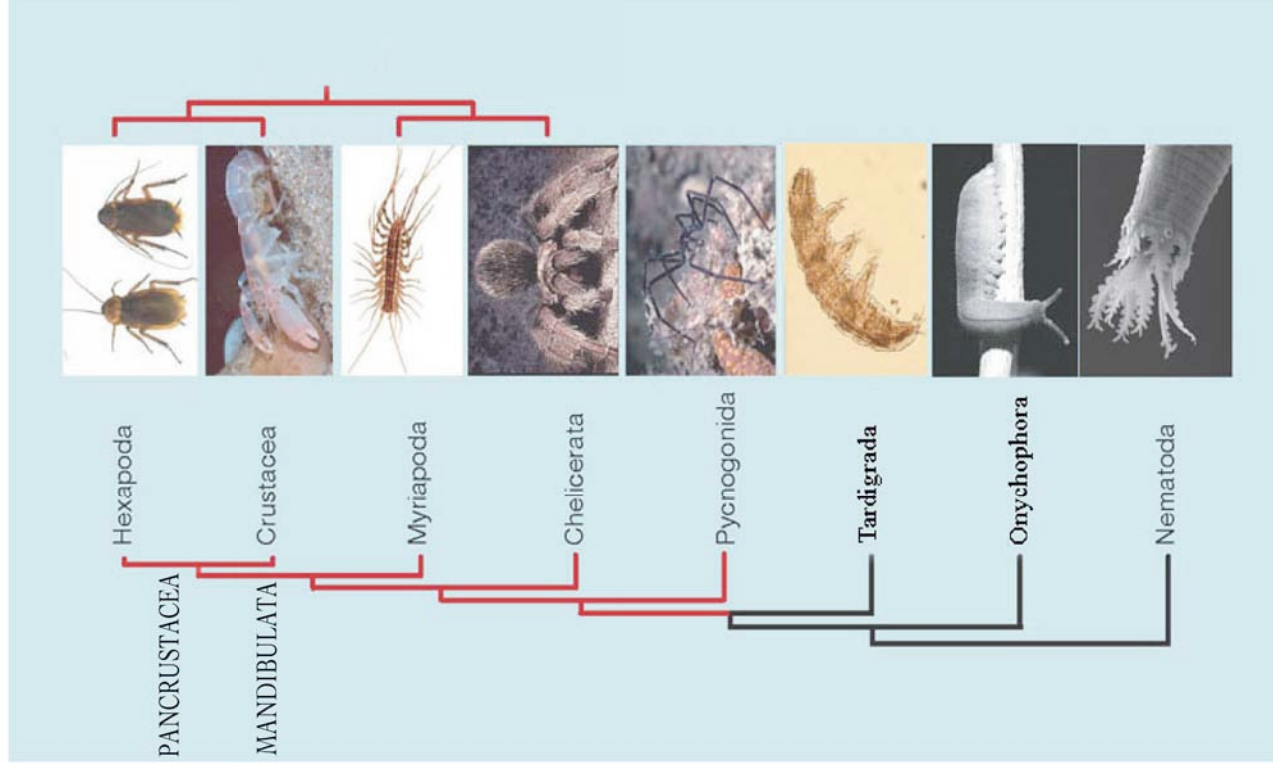
An appreciation of arthropod phylogeny is fundamental to gaining a correct understanding of evolution in arthropod body plans, and yet certain relationships between and within the major extant Arthropod sub-phyla - Chelicerata, Myriapoda, Crustacea and Hexapoda/Insecta – have been a long standing topic of dispute, still partly unresolved.

Accepted relationships

Consensus now rests securely on the monophyly of the Arthropoda as a whole, as well as of each sub-phylum Chelicerata, Myriapoda, Crustacea and Insecta (Giribet et al. 1999; Wheeler et al. 2004). Among diverse phyla of the ‘Ecdysozoa’ – a major protostome clade comprising moulting animals – the Onychophora and Tardigrada are consistently retrieved as closest outgroups to the arthropods, useful to root arthropod phylogenies and allow inferences about directions of evolutionary modification (c.f. Figure 1.3.1) (Cracraft and Donoghue 2004; Giribet et al. 1996).

It is widely accepted that Crustacea + Insecta form a clade separate from the Myriapoda and Chelicerata, termed the ‘Pancrustacea’ or ‘Tetraconata’ (c.f. Figure 1.3.1). The Tetraconata were named after shared, quadri-crystalline ommatidial structure, other synapomorphies including optic neuropil arrangement and mode of neurogenesis. However, there are proponents of a conflicting

(i)



(ii)

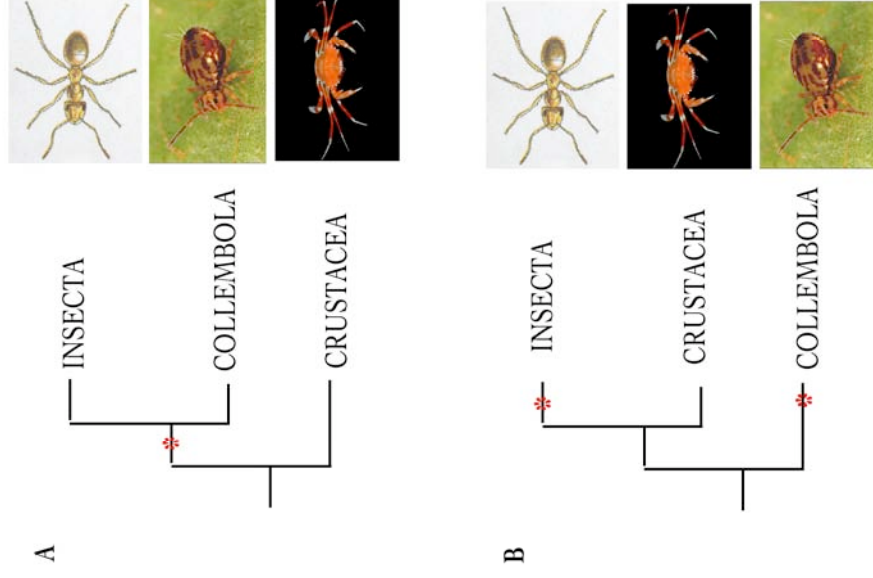


Figure 1.3.1 Arthropod phylogeny, indicating major current points of debate:
(i) Based on Fig. in Blaxter (2001), with updated branching order of outgroup taxa based on data in Cracraft & Donoghue (2004), Dunlop & Arango (2005). On the left, the favoured topology, with chelicerates basal to the Mandibulata (e.g. Giribet et al, 2001). On the right, Pancrustacea separated from sister taxa Chelicerata + Myriapoda, with no indication as to placement of Pycnogonida (e.g. Hwang et al, 2001; Kusche & Burmeister, 2001).
(ii) A: Traditional interpretation of hexapod phylogeny, Collembola basal to the true insects. B: Mitochondrial DNA phylogeny (Nardi et al, 2005), inferring evolution of a hexapod bodyplan (**red asterisk**) twice, as Collembola branch basal to the group Crustacea + Insecta within Pancrustacea.

Myriapoda + Insecta clade, originally termed the 'Atelocerata' or 'Tracheata', based on unbranched trunk limbs and common tracheal structure. Tracheae could have evolved convergently in response to terrestriation, but further molecular and especially fossil morphological data is required to resolve this issue completely (Cracraft and Donoghue 2004; Telford and Thomas 1995; Wheeler et al. 2004). Recent molecular phylogenies using *Hox* genes and combined datasets discredit the Atelocerata hypothesis and imply, furthermore, that insects originated from within the Crustacea rather than as a strict sister group (Telford and Thomas 1995).

Debated relationships

Most but not all recent analyses have converged on a robust clade 'Mandibulata', grouping myriapods + crustaceans + insects, a group originally proposed on the basis of shared possession of a mandible; a sclerotised mouthpart formed mainly from the coxa due to secondary proximo-distal segment reduction (Figure 1.3.1) (Telford and Thomas 1995). Molecular evidence for a sister group relationship between chelicerates and myriapods has been presented, including a shared α -helix deletion in spider and diplopod haemocyanins and gene order within mitochondrial genomes (Hwang et al. 2001; Kusche and Burmester 2001). Furthermore, developmental genetic analysis of neural cell cluster recruitment and invagination during spider and diplopod neurogenesis reveals remarkable commonalities, contrasting with equally notable conservation in modes of neurogenesis shared between crustaceans and insects (Harzsch 2003; Mittmann and Scholtz 2003; Osorio et al. 1997; Stollewerk et al. 2003a; Stollewerk et al. 2003b). However, in a large study combining gross morphology, ultrastructure and developmental traits with molecular data (2 mitochondrial genes, 3 nuclear and 3 ribosomal protein-encoding genes), chelicerates resolve as basal arthropods, the myriapods branching as an intermediate group between chelicerates and the more derived crustacea + insects (Blaxter 2001). Chelicerates also branched basal to mandibulates in a study combining morphological traits with available gene sequence data (28S + 18S ribosomal DNA), corrected for taxonomic bias by including both extinct and extant taxa (Wheeler et al. 2004). This topology is further supported by the extent of specialisation in *Hox* gene expression along the antero-posterior (A-P) axis: extensive overlap is observed in Chelicerate *Hox* gene expression domains, relative to an intermediate condition in myriapods and more refined, segment-specific deployment in crustaceans and insects (Abzhanov et al. 1999; Hughes and Kaufman 2001; Pavlopoulos and Averof 2002). In addition, the *Hox* gene *zen* is expressed in a typical, restricted A/P domain in chelicerates, has diverged significantly to acquire a new role in extra-embryonic development in insects, and appears to retain both ancestral *Hox*-like and derived roles in myriapods (Hughes and Kaufman 2001; Telford and Thomas 1998b).

Within the Pancrustacea, the placement of Collembola, with 3 pairs of thoracic and 3 pairs of modified abdominal appendages, is a point of new controversy: recent mitochondrial DNA evidence places Collembola basal to true insects + crustaceans, rather than in the previously accepted position as basal Insecta(Nardi et al. 2003a; Nardi et al. 2003b). If proved correct, the hexapod body plan - defined by a tripartite body and six thoracic walking legs - would no longer be monophyletic, having evolved twice independently; in the Collembolan lineage and in the lineage leading from some form of non-hexapod crustacean ancestor to true insects (see Figure 1.3.1).

Lastly, in their relationship to chelicerates, the enigmatic pycnogonids, with 8 or (more rarely) 10 locomotory limbs, are of uncertain phylogenetic affinity(Coddington et al. 2004). Recent mitochondrial, nuclear and ribosomal-protein encoding sequence analysis supports the classical hypothesis of pycnogonids as basal chelicerates, whereas combined morphological and molecular analyses indicate they are either sister to chelicerates, or sister to all arthropods(Blaxter 2001; Briggs and Collins 1988; Coddington et al. 2004; Dunlop and Arango 2005). The morphology of a spectacularly preserved fossil larval pycnogonid, discovered in Upper Cambrian 'Orsten' phosphatic deposits, supports a Pycnogonida + Euchelicerata (true chelicerates) relationship, basal to all other arthropods(Blaxter 2001; Waloszek and Dunlop 2002).

Chelicerate comparisons

Given the position of chelicerates as basally branching arthropods, external to the Pancrustacea and likely external to the Mandibulata, the group serves to broaden the phylogenetic reach of possible evolutionary scenarios arising from developmental genetic comparisons with derived arthropods such as the model insect *Drosophila*(Simonnet 2005). In addition, within the sub-phylum the Chelicerata display an impressive range of variation in form (detailed in 1.3.2), lending them to examination of morphological innovation across orders down to the level of cross-species comparison(Anderson 1999; Cloudsley-Thompson 1958; Coddington et al. 2004; Foelix 1996; Selden et al. 2005).

1.3.2 The spider mite: a Chelicerate model for evo-devo?

Phylogenetic placement

The Chelicerata (Figure 1.3.2a) are sub-divided into two major 'Euchelicerate' Classes, Merostomata and Arachnida, with Pycnogona included as a minor Class in some phylogenies (Anderson 1973; Anderson 1999; Brusca and Brusca 2003; Wheeler and Hayashi 1998). Merostomata comprise a single Order, the Xiphosura (horse-shoe crabs), and have a mid-ventral mouth and thick prosomal carapace. The Arachnida comprise 11 Orders, of which Scorpiones are generally presumed basal, and Acari (mites + ticks) are by far the most speciose, followed by Araneae (spiders); other less

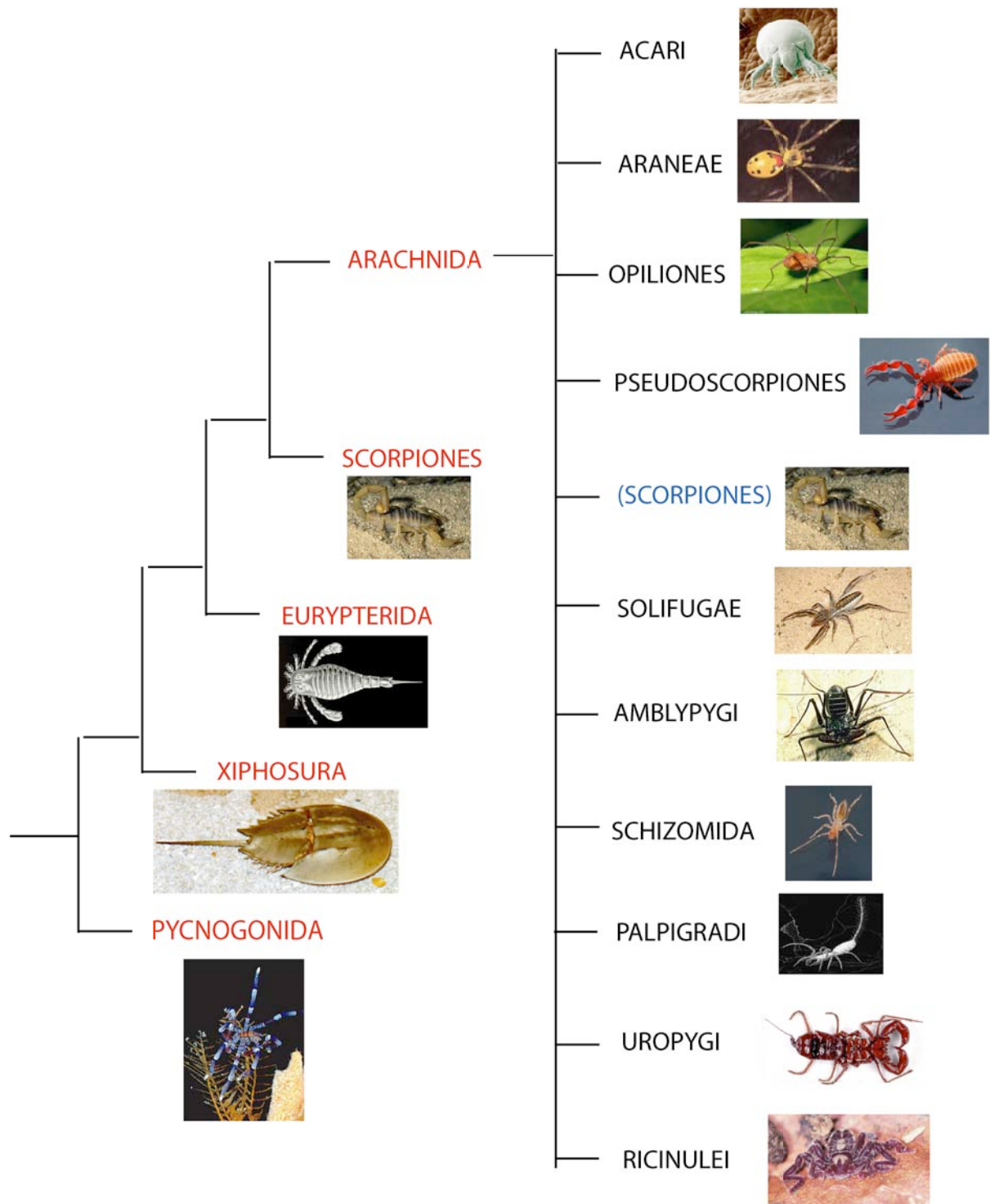


Figure 1.3.2a Phylogeny of the Chelicerata, showing all Orders within the Arachnida. Eurypterids (drawing by E. Haeckel) are the only extinct group included. Phylogeny based on Coddington & Wheeler (2004) *in* Cracraft & Donoghue (2004), and Simonnet (2005), who points out lack of resolution with respect to Scorpiones.

known but remarkable Orders include Opiliones, Pseudoscorpiones, Solifugae, Amblypygi, Schizomida, Palpigradi, Uropygi and Ricinulei (Barnes et al. 2001; Coddington et al. 2004). Although Xiphosuran embryology and much arachnid embryology has been described morphologically (e.g. Anderson, 1973; Farley 2005, 2001, 1998), developmental genetic studies have had a narrow taxonomic sampling, focussing on spiders *Cupiennius salei*, *Achaeareanea tepidariorum* and *Steatoda triangulosa*, acarid *Archegozetes longesitosus*, scorpion *Paruroctonus mesaensis* and Xiphosuran *Limulus polyphemus* (Abzhanov and Kaufman 2000c; Abzhanov et al. 1999; Anderson 1973; Cartwright et al. 1993; Farley 1998; Farley 2001; Farley 2005; Mittmann 2002; Mittmann and Scholtz 2001; Mittmann and Scholtz 2003; Popadic and Nagy 2001; Telford and Thomas 1998a; Telford and Thomas 1998b; Thomas and Telford 1999). Most recently, work on the spider *Tegenaria saeva*, acarid *Tetranychus urticae* and scorpion *Euscorpius flavicaudius* has contributed more data to, and strengthened, the burgeoning field of comparative evo-devo (Dearden et al. 2003; Dearden et al. 2002; Simonnet et al. 2005a; Simonnet et al. 2004; Simonnet et al. 2005b). I propose to further our understanding of development in the two-spotted spider mite *Tetranychus urticae*, building on limited but interesting work that has been done in this species.

Acari (Figure 1.3.2b) are divided into Opilioacariformes, Parasitiformes (liquid feeders) and Acariformes, the latter subdivided into two major branches, Trombidiformes and Sarcoptiformes – primarily distinguishable according gnathosomal morphology (Brusca and Brusca 2003; Coddington et al. 2004). The only other acarid studied so far in evo-devo is *Archegozetes longesitosus* which is a sarcoptiform, whereas *Tetranychus urticae* is a trombidiform (Telford and Thomas 1998a; Thomas and Telford 1999). Within Trombidiformes, *Tetranychus urticae* belongs to the derived family Tetranychoidae within the Prostigmata (Figure 1.3.2b), showing extreme gnathosomal modification as a secondary adaptation to phytophagy (Boudreaux 1963). Limitations can be encountered when interpreting ‘ancestral’ states using a highly derived species as it could have secondarily lost or modified many of its characters, but it is still equally plausible for a more basal species to possess derived characters: a consideration of specific ecological conditions or selective pressures may help to accommodate or reduce this problem. Indeed, the ‘model’ Dipteran *Drosophila melanogaster* is an insect with many derived characters, and although this limits the scope of comparative conclusions - a point not generally articulated by more ‘Drosophilocentrist’ authors - its mode of life as an opportunist feeder requiring very rapid development is illuminating: for example, it affects the topological distribution of gene products that pattern the specialised, two-dimensional larval ‘imaginal discs’ that telescope outwards to form adult limbs (Simonnet 2005).

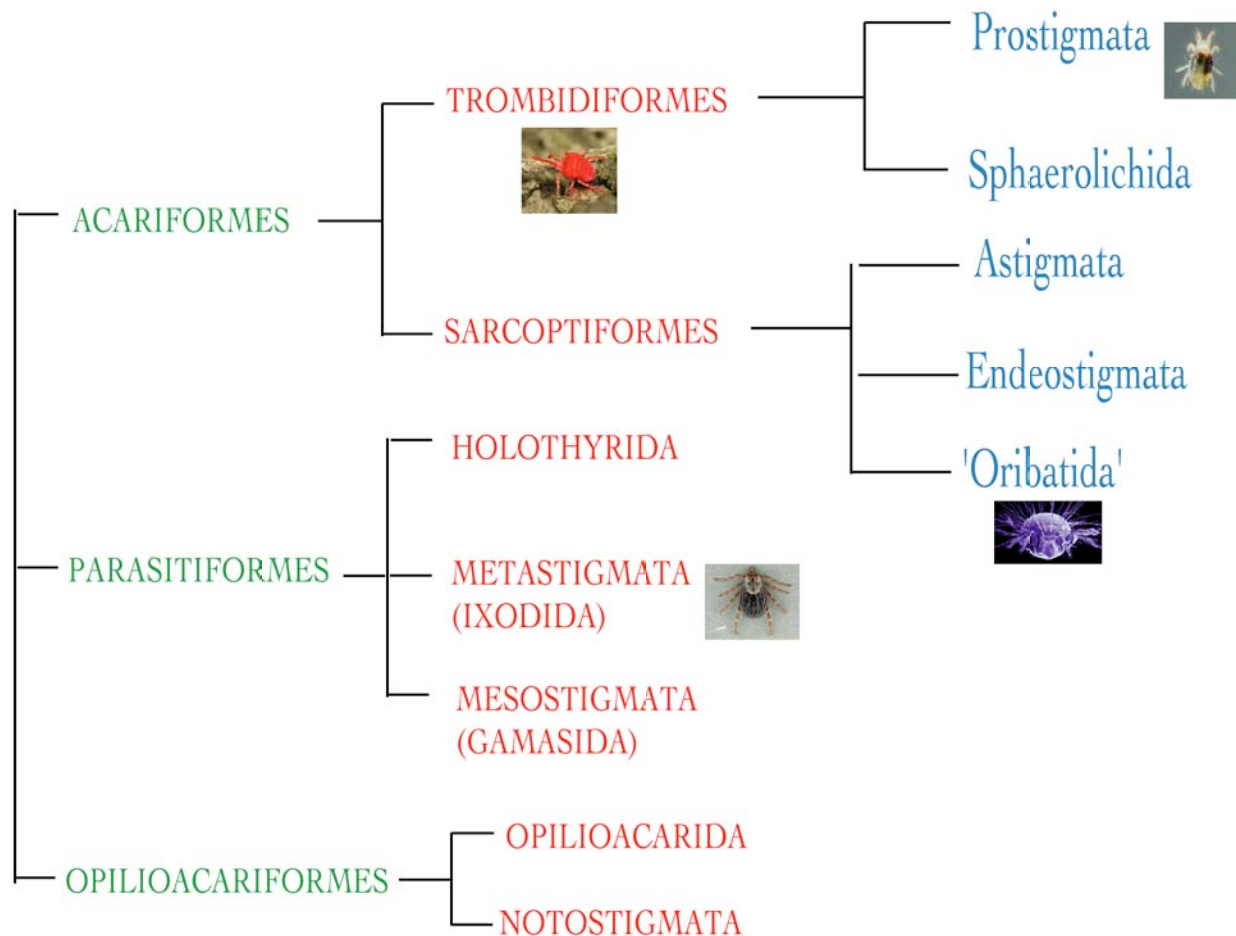


Figure 1.3.2b Phylogeny of the Acarida. **Green:** Super-Orders, **Red:** Orders, **Blue:** Sub-Orders. Images included for groups studied so far regarding genomics/ESTs (e.g. ixodid *Dermacentor*, prostigmatan *Tetranychus urticae*), physiology (e.g. Ixodid *Amblyomma americanum*) and evolutionary developmental biology (e.g. oribatid *Archegozetes longisetosus*, Prostigmatan *Tetranychus urticae*).

Information for phylogeny derived from Coddington & Wheeler (2004) and Woolley (1964).

Experimental suitability

Tetranychus urticae is a pest mite species, phytophagous on agricultural and horticultural plants. Large populations of spider mites can be cultured very easily on young bean plants (e.g. *Phaseolus* spp.), and although they tolerate a wide temperature range, will reproduce most rapidly between 25-37°C (Collyer 1998). Reproduction is either sexual or by female parthenogenesis: females lay eggs singly on the surface of leaves, up to ~100 eggs in a lifetime if fertilised or 50-60 eggs in a lifetime if unfertilised (i.e. parthenogenetic) (Florida 2005). The life cycle is rapid, requiring 6-20 days for embryogenesis, then larval, nymphal and mature, reproductive adult moults to be completed, the rate dependent on temperature (see Chapter II for details of *Tetranychus urticae* development) (Bynum and Porter 2006). The fastest rates of embryonic and post-embryonic development occur at optimal temperatures of 25°C or above, and in low humidity (Bynum and Porter 2006).

Spider mite extra-embryonic and outer egg membranes (i.e. vitelline membrane and chorion layer) are highly impermeable but can be removed, rendering embryos amenable to standard techniques of arthropod fixation and assays for developmentally relevant gene expression or protein activity domains. In spite of impermeability, egg membranes are nearly transparent, allowing direct *in vivo* observation of nuclear divisions during early zygotic cleavage and blastoderm formation, and general morphology as embryogenesis progresses. For example, a lateral or transverse light source creates contrast between yolk and cytoplasmic 'haloes' that surround cleavage nuclei, permitting detailed, timed observation of cell movement, karyo- and blasto-kinesis during different mitoses (see Figure 2.1.1a, Chapter II).

Estimates of genome size for *Tetranychus urticae* approximate to 75Mbp (0.75×10^8 bp), smaller in size than that of model arthropod *Drosophila melanogaster* (180Mbp) and comparable to the *Caenorhabditis elegans* (100Mbp) genome (Davidson 2001). The small genome size makes *Tetranychus urticae* one of the more suitable species to consider for full genome sequencing; the Joint Genome Institute run by the University of California for the US Department of Energy (DOE-JGI) are currently preparing to sequence the spider mite. Genomic data will augment sequence data already retrieved from an EST project carried out at Agriculture Canada in association with the University of Western Ontario and a limited number of candidate gene fragments successfully obtained by evo-devo research into spider mite orthologs of genes involved in germ cell specification and segmentation (Dearden et al. 2003; Dearden et al. 2002). A complete database for the *Tetranychus* genome, transcriptome and proteome will be invaluable in providing chelicerate data-points for use in phylogenetic and developmental genetic comparisons.

The haplo-diploid reproductive mode of *Tetranychus urticae* makes the species tenable for F1 hybrid mutant screening and establishment of transgenesis as a means of determining putative gene function

via particular mutations. In addition, RNA-interference (RNAi) mediated gene silencing methods have been tested in the form of embryonic and parental (maternal) double-stranded RNA (dsRNA) injection experiments in various animals; it has been shown that short dsRNA fragments synthesised to complement a specific gene can induce *in vivo* activity of innate cellular-genetic machinery that causes gene silencing, similar to an immune response (c.f. section 1.1.2, footnote 7). Limited positive results testing this powerful new genetic tool indicate possible potential for applying RNAi to generate targeted gene knockouts and to characterise highly specific gene loss-of-function phenotypes(Khila 2006b).

1.4 Aims and Structure of this Thesis

1.4.1 Thesis Aims

1. To describe embryonic and post-embryonic development in the spider mite *Tetranychus urticae*, with a particular emphasis on limb formation.
2. To develop and demonstrate a working protocol for whole mount *in situ* hybridisation and antibody staining in *Tetranychus*.
3. To clone the *Tetranychus Dll* homolog and to verify its conserved expression during limb development; transcription is predicted from limb specification onwards.
4. To clone *Tetranychus* orthologs for candidate genes potentially involved in specification of limb primordia *via* direct or indirect regulation of *Dll* – i.e. the genes *wingless*, *engrailed*, *dp*, *EGFR*, *Ubx*, *abdA* and *Sp*.
5. To determine mRNA and/or protein expression patterns for those candidate *Dll*-regulatory genes during embryogenesis, with particular attention to early limb development.
6. To compare sequence data and expression profiles of the *Tetranychus* candidate genes with data available for orthologs of these genes in other arthropods and in more distantly related species where relevant, thereby to assess the possible nature of conservation or divergence in gene regulatory networks operating during arthropod limb specification, consistently marked by *Distal-less* expression.
7. To assess the validity and practicality of using *Tetranychus urticae* as a model species for evolution and development studies.

1.4.2 Thesis Structure

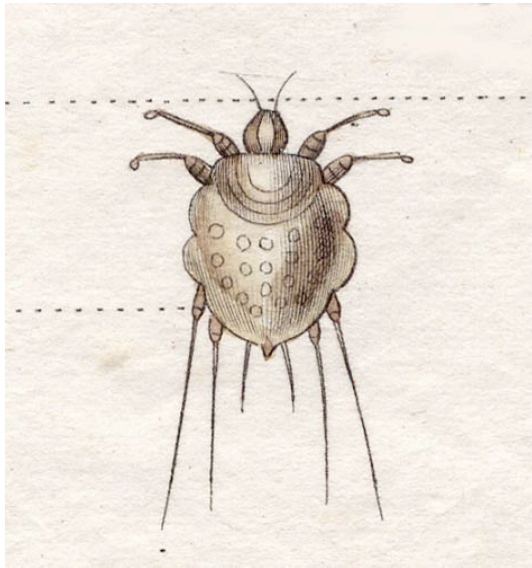
I have introduced *Tetranychus urticae* as a chelicerate species through which to examine potential evolutionary changes in *Dll* regulation that relate to appendage development, using the Arthropoda as a model phylum for a comparative 'evo-devo' approach. Chapter II progresses to give a detailed description of embryonic and post-embryonic development of *Tetranychus urticae*, the two spotted spider mite, as the basis for interpretation of mRNA and protein expression domains revealed by whole mount *in situ* hybridisation and antibody detection. In Chapter III, I confirm the conserved nature of *Dll* expression in *Tetranychus urticae*, and briefly consider the existence and role of *Sp* genes in activating a genetic program for limb specification. Chapters IV and V document and discuss results pertaining to *Tetranychus* orthologs of genes that establish generic antero-posterior and dorso-ventral co-ordinates for segmental limb primordia in *Drosophila melanogaster*. Chapter VI addresses the potential role of *BX-C* genes in regulating early *Dll* in the limbless spider mite opisthosoma, presenting and discussing results pertaining to *Tetranychus Ultrabithorax* and *abdominal-B*. In Chapter VII (Discussion), I bring together issues arising from Chapters III-VI, including interpretations of genetic interactions based on expression data and limited inferences regarding conservation and divergence of the early *Dll* regulatory network throughout the Arthropoda. I briefly discuss the issue of deep homology in Bilaterian limb developmental genetic networks, and make an assessment of future work that would be required to determine further information about mechanisms of body plan evolution given the nature of *Dll* regulation in *Tetranychus urticae*. Finally, I assess the feasibility of using *Tetranychus urticae* as a new model system for developmental genetic studies, and given problems encountered, make suggestions regarding other possible chelicerate model organisms to pursue. Materials and Methods relevant to protocols and techniques used during experimental work and analysis are given in Chapter VIII, followed by a full bibliography and Appendices. Figures are interleaved throughout the text, singly or collated in order directly after the text page in which each Figure is first referred to.

CHAPTER II

OVERVIEW OF DEVELOPMENT IN THE TWO-SPOTTED SPIDER MITE

TETRANYCHUS URTICAE KOCH 1836

(ACARIDA: PROSTIGMATA: TETRANYCHIDAE)



Mite de la gale

Introduction to Chapter II

Acarid embryogenesis is primarily known from work on parasitiform ticks, which are basal within the Acariformes, and from recent work on the euoribatid mite *Archegozetes longisetosus*, a divergent species within the Sarcoptiformes (Anderson 1973; Thomas and Telford 1999). Spider mites such as *Tetranychus urticae* Koch belong to the opposing major acariform clade, the Trombidiformes, and development, both embryonic and post-embryonic, has not been investigated in detail in any of the 122 species in the genus *Tetranychus* Darfour (Anderson 1999; Brusca and Brusca 2003; Brusca and Brusca 1990; Flechtmann and Knihinicki 2002; Koch 1836). This Chapter constitutes an examination of *Tetranychus urticae* development, from first zygotic cleavage to reproductive adult. Whole mount examinations were carried out at various developmental stages, employing light microscopy, nuclear fluorescence labelling (with DAPI), mitotic cell labelling (with α -pHistone-III), time lapse microscopy and scanning electron microscopy, in addition to some fluorescent dye injections when defining the nature of early cleavage. The nature of development revealed for this spider mite adds to our appreciation of the diversity in mechanisms of embryogenesis and pattern formation within acarids and the Chelicerates in general.

Overview of *Tetranychus urticae* development

A low cuboidal blastoderm is formed after 9 cleavage divisions, subsequently proliferating on one side to form a germinal disc between 10-12hrs after egg laying (AEL). Ingression is observed at one end of the germdisc, representing the site of gastrulation: mesoderm cells invaginate and migrate anteriorly beneath a germband that extends from a posterior growth zone around the yolk until posterior and anterior extremes are opposed on the dorsal surface. Segmentation grooves appear first at the posterior, indicative of an unusual initial posterior to anterior pattern of differentiation, followed by near-simultaneous appearance of ventro-lateral limb primordia (cheliceral to L3 walking limb buds), at 21-22hrs AEL. Limb buds grow and elongate as the medial ventral ridge closes, and further appendage morphogenesis and sensory organogenesis proceeds as the germband contracts and undergoes dorsal closure (45-50hrs AEL). The first moult occurs within the egg, giving a 'pre-larva' with pigmented eyes visible through the chorion. Hexapod larvae hatch at 70-75hrs AEL, L4 limb buds developing and becoming functional only from the next, pronymphal moult. Moulting occurs approximately every 24hrs, pronymphs followed by deutonymphs then sexually mature, phytophagous adults. Females lay either fertilised or unfertilised eggs of approx. 75 μ m diameter, on the surfaces of plant leaves or on webbing spun over heavily infested host plants.

2.1 Embryonic development

2.1.1 Cleavage and blastoderm formation

Females lay eggs singly, either on the lower surface of a host plant leaf, or attached to webbing that is formed on heavily infected leaves. Eggs are spherical, with an average diameter of 75 μm ¹² (see Table 2.1.1). Initially, yolk is evenly distributed throughout the egg, but mitotic cleavage divisions generate a centrolecithal structure; an outer cellular blastoderm separated from an inner, virtually anucleate yolk mass (see Figures 2.1.1a, 2.1.1b). In more primitive acarids, the egg is surrounded by a vitelline membrane and bilayered chorion, but *Tetranychus* is more typical of small, derived acarids in having a vitelline membrane and chorion reduced to a single layer (Anderson 1973; Boudreaux 1963; Brusca and Brusca 1990).

Table 2.1.1 Spider mite egg diameter measurements

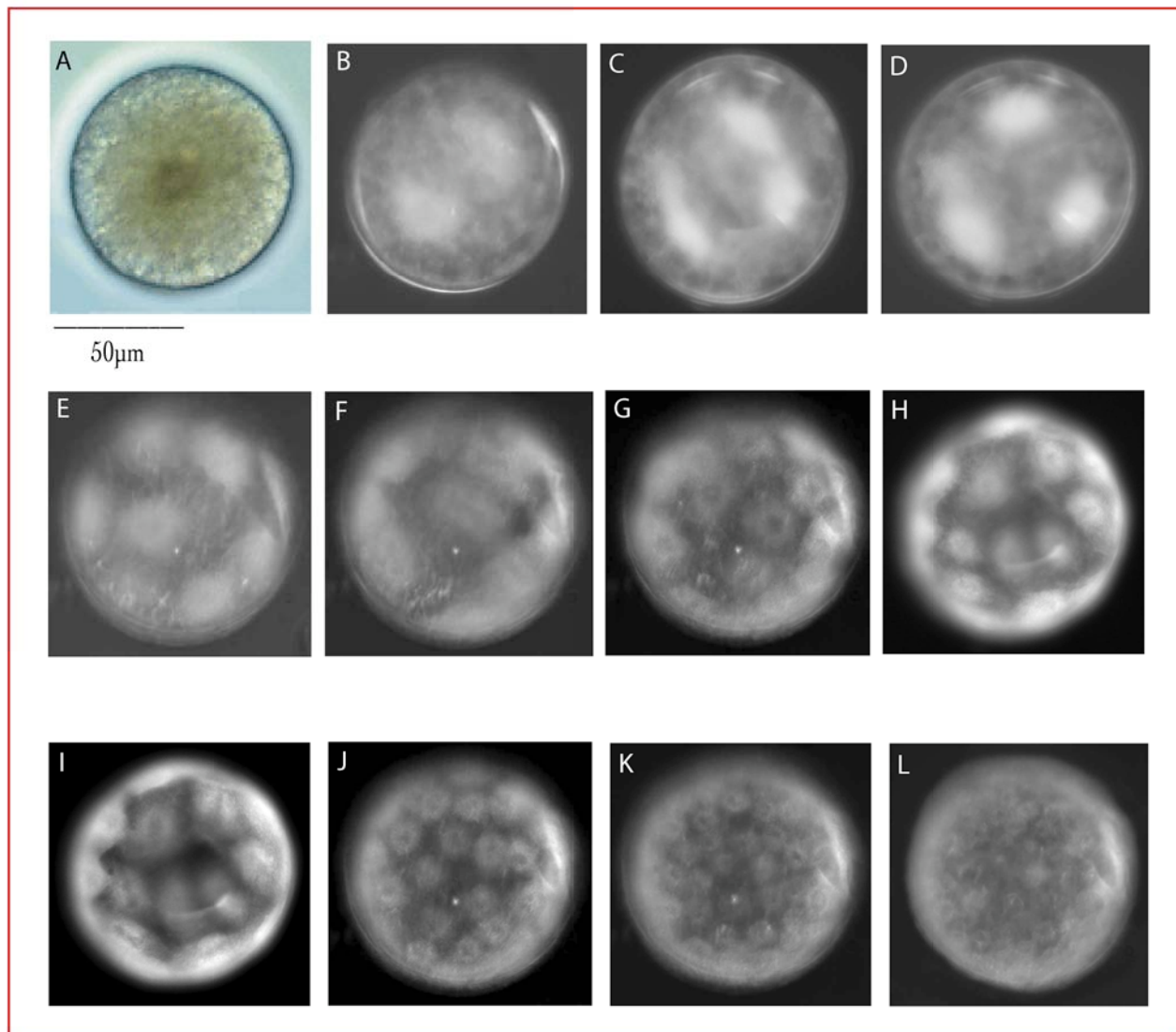
EMBRYO DEVELOPMENT STAGE	AVERAGE EGG DIAMETER (μm)
Blastoderm	72.0
Germband	70.1
Limb buds - early outgrowth	79.6
Late limb growth and jointing	72.0
Prelarva	76.0
Total	73.9

i) Early cleavage

Nuclear chromatin condenses at the centre of the egg between 0-1 hr AEL (After Egg Laying), indicating exit from G1 phase or interphase, in which chromatin is diffusely distributed, and progression to G2 and prophase of zygotic mitosis (Alberts et al. 1994; Gilbert 2000). The first mitotic cleavage is total (i.e. holoblastic); two daughter cells of virtually equal size are produced, nuclei and cytoplasm separated by a membrane that completely transects the egg. As described in Anderson (1973), cytoplasm surrounding the nuclei is denser than the near-transparent yolk, so that

¹² Embryo measurements were taken after membrane removal, hence the overall average of 73.9 μm in Table 2.1.1 was corrected to 75 μm for intact, osmotically balanced eggs.

(i)



(ii)

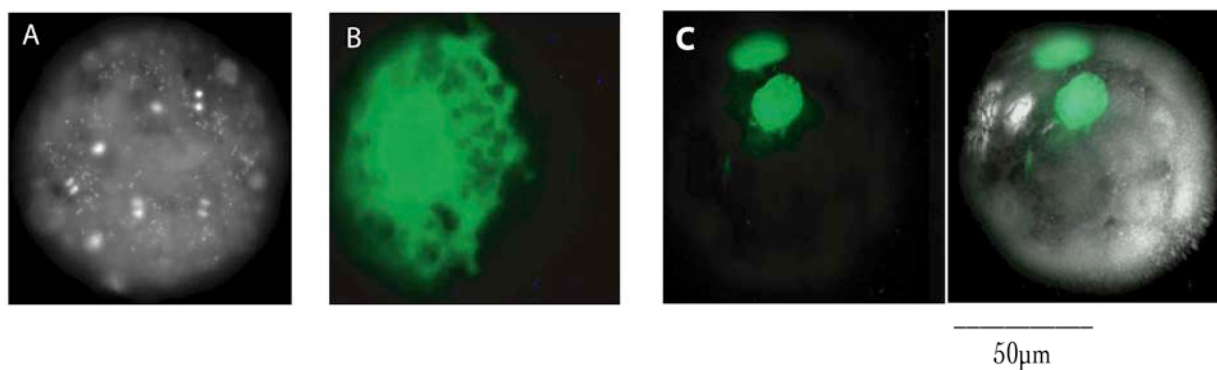


Figure 2.1.1a

(i) Cleavage divisions from zygote to blastoderm stage embryo. A: 1 cell, B: 2 cells, C: 2nd mitosis, D: 4 cells, E: 8 cells, F: 4th mitosis, G: 16 cells, H: 32 cells, I: 6th mitosis, J: 64 cells, K: 128 cells, L: 256 cell blastoderm, ~7hrs AEL.

(ii) A: DAPI labelled nuclei at 5th mitosis. Evidence for total cleavage; B: FITC fluorescent dye injection into 2-cell embryo. C: inheritance of FITC dye in daughter cells of 16-32 cell embryo. [1-cell image from Dearden et al (2002).]

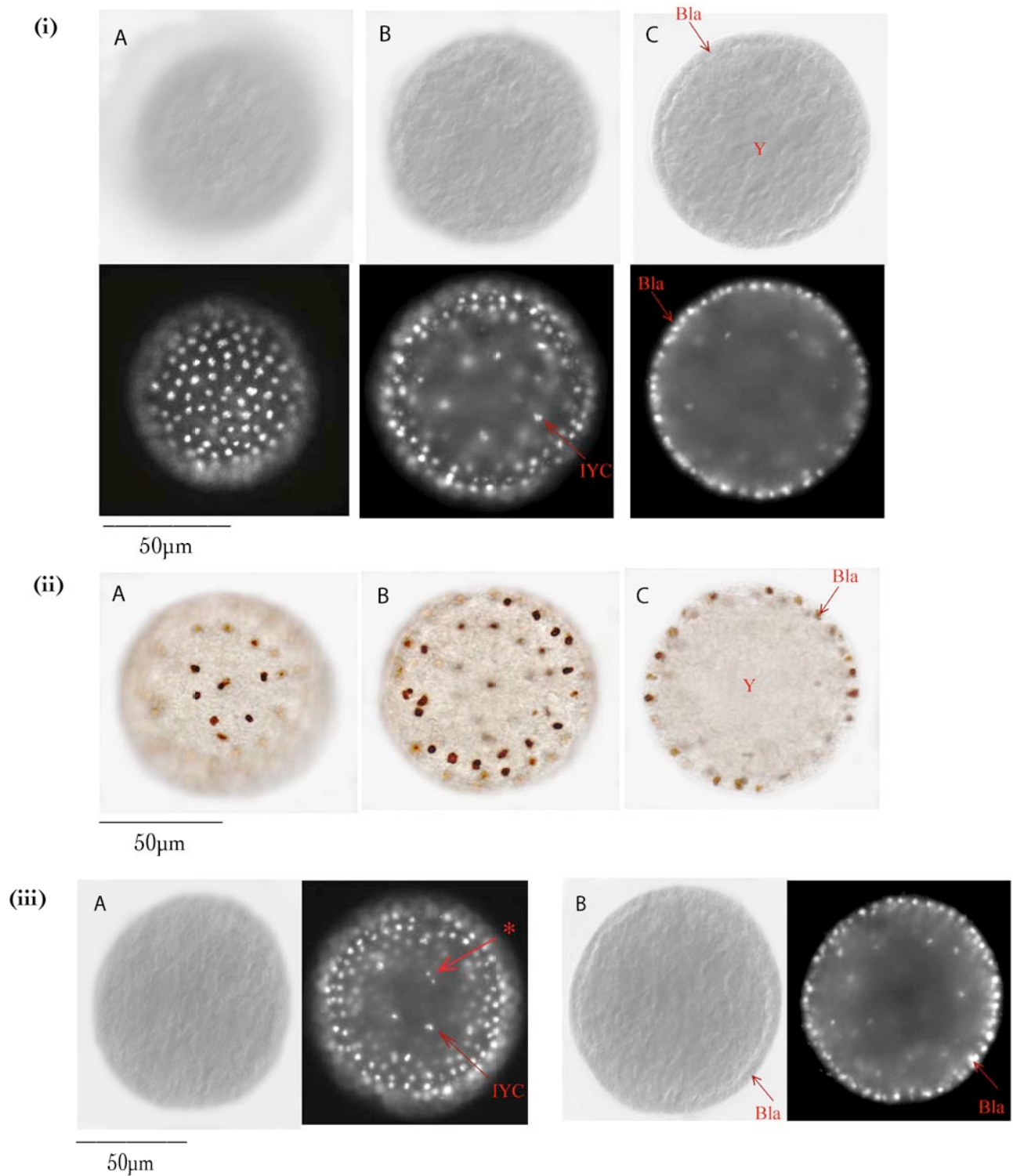


Figure 2.1.1b *Tetranychus urticae* blastoderm stage of embryogenesis.

(i) Blastoderm stage embryo, sections from surface (A) to equatorial (B); DIC and DAPI.

(ii) Anti-phospho-Histone-III labelling of blastoderm nuclei, 3 sections.

(iii) Later blastoderm embryo, cells appear slightly denser on one side (right). Near-surface (A) and equatorial (B) sections, DIC and DAPI.

Bla - blastomere cell layer; Y - yolk; IYC - inner yolk cell; * - recently divided cells

white cytoplasmic 'haloes' are visible around darker, condensed nuclear material prior to each cell division (Figure 2.1.1a,i) (Anderson 1973).

Mitotic spindles are formed approximately perpendicular to that of any adjacent sister blastomere, and also perpendicular to the immediately previous parent-daughter cleavage plane. This consistent relative orientation at each division facilitates energid movement to the periphery of the egg, with associated displacement of yolk towards the interior (Anderson 1973). Complete cellularisation occurs between every mitotic cycle: fluorescent dye (e.g. FITC) injected into one cell of a 2-cell embryo remains localised to the same cell, clearly demonstrating the selectively permeable membrane barrier present (Figure 2.1.1a,ii; see also (Dearden et al. 2002)). Total cleavage was further shown by tracing the inheritance of injected dye from a specific parent cell into its daughters (Figure 2.1.1a,ii). The first 5 cycles of cell division are clearly well synchronised, but synchronicity decreases from cycle 5 (32-cell stage) onwards. Eight cleavage cycles take place in total, each lasting approximately 45' from mitosis to the next G1 interphase (i.e. membranes newly-formed, blastokinesis complete) (Alberts et al. 1994).

ii) Blastoderm stage

The cleavage process results, after ~7hrs, in formation of a blastoderm composed of an outer layer of 256 near-cuboidal blastomeres, approximately 7.5µm thick apical to basal, and an inner unitary yolk mass - as is typical for acarids (Anderson 1973). Figure 2.1.1b shows blastoderm structure by means of DIC imaging, nuclear fluorescence (DAPI labelling) and a phospho-HistoneIII antibody that targets dividing centromeres. Between 7 and 9hrs AEL the blastomere layer undergoes slight attenuation to ~4.5µm thick, and cell density appears to increase particularly on one hemisphere of the blastoderm, foreshadowing the germdisc (Figure 2.1.1b,iii).

iii) Internal non-blastodermal cells

I observed internal 'vitellophage' nuclei scattered within the yolk as the blastoderm forms and during its subsequent development (see Figure 2.1.1b,i for example). These 'intra-yolk' cells have been proposed to originate primarily from the compact group of pre-mesodermal cells that proliferates later on, just beneath the developing germdisc, and secondarily direct from the blastoderm (Anderson 1973). In *Tetranychus*, at least during cleavage and initial blastoderm formation, these cells may be either remnant cleavage energids capable of autonomous subsequent mitoses, or, in partial agreement with previous ideas, that they derive from the blastoderm.

The fate of intra-yolk cells has not been clearly shown within Chelicerata, neither generally nor specifically in the Acarida. Dearden et al. (2003) propose a germ cell fate for early *Tetranychus urticae*

non-blastodermal cells, but i) see comments on that article in Chapter VII: Discussion, and ii) a comparison of embryonic development in primitive acarids and basal chelicerates such as *Xiphosura* suggests that these cells develop as true vitellophages or acquire an alternative fate in midgut specification (Anderson 1973; Dearden et al. 2003). Observations with respect to this enigmatic cell population in *Tetranychus* are therefore made in subsequent sections of this chapter with the aim of clarifying this issue.

2.1.2 Germdisc stage

i) Germdisc accumulation

From 10-12hrs AEL a dense cluster of germdisc cells becomes clearer, a convex internal boundary delimiting an oval-shaped thickening that occupies 25-30% of the total egg circumference, cell density decreasing away from the disc.

Few cells are present within the yolk during germdisc accumulation. By DAPI nuclear labelling I observed their asynchronous division (Figure 2.1.1b,iiiA), and in some cases a slightly higher cell density close to the accumulating germdisc (Figure 2.1.2). Independent, asynchronous cell division within the yolk suggests a degree of autonomy as remnant cleavage cells or as derivatives from blastodermal cell lineages.

ii) Germdisc formation and gastrulation

Between 12-15hrs AEL the germdisc attains a more defined morphology, dynamic yet distinct from the remaining attenuated blastoderm, and preliminary cellular ingression occurs at a gastrulation groove, initiating formation of the triploblastic germ layers (ectoderm, mesoderm, endoderm).

A circular germdisc appears, representing a discrete body of cells relative to the blastomeres, occupying ~30% of the circumference. Viewed in cross-section, the disc is thickened and, at its centre, projects internally by ~9µm (12% total egg diameter), with mild swelling also noticeable on the outer surface (Figure 2.1.2, i). At this point the blastodermal layer is attenuated to ~3.75µm; half the thickness of the 256-cell stage blastoderm.

Shortly after the germdisc becomes clearly defined, its circular outline is transformed to a more oval shape (Figure 2.1.2,ii-iv). Coincident with this change in form, an invagination groove becomes visible, corresponding to Anderson's 'gastral groove' (Anderson 1973). Active invagination is indicated by cellular ingression when viewed in cross-section and by accumulation of a cluster of

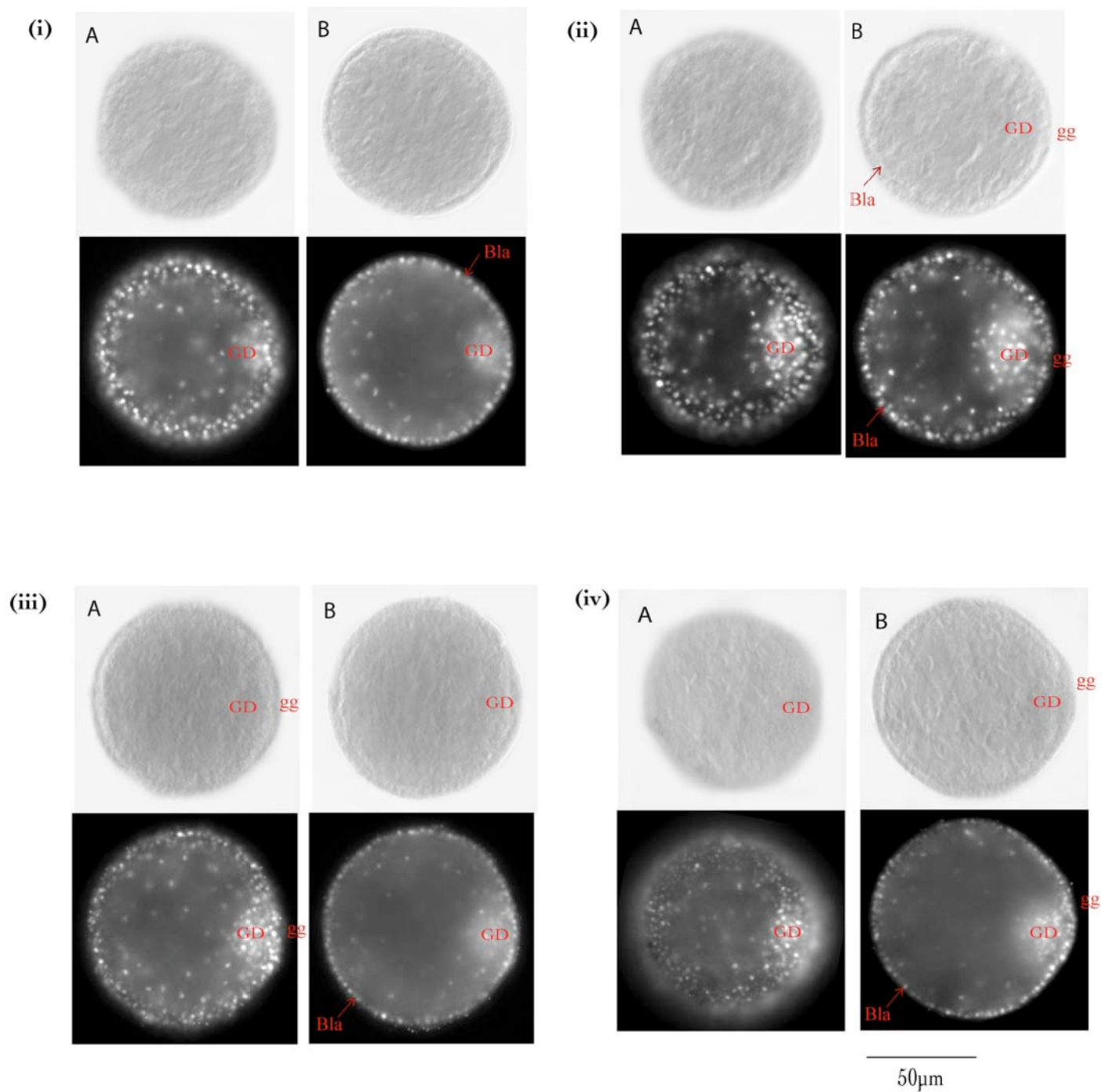


Figure 2.1.2

(i) Early germdisc; circular. (ii) - (iv) Later germdiscs; oval form, slight protrusion from peripheral surface and early ingress groove at posterior. Separation between ectodermal germ disc swelling and ingressed, putative endodermal cells. A: transverse, B: equatorial. DIC above, DAPI below.

Bla - blastomere cell layer; GD - germ disc; gg - gastral groove ingress site

relatively large cells positioned beneath, but separate from, the germdisc (Figure 2.1.2,ii-iv). The germdisc and associated interior cell mass are $\sim 10.4\mu\text{m}$ thick (14% of diameter) at this stage. The inner cell mass may represent primordial midgut endoderm, previously termed the 'posterior midgut rudiment' in other acarids (e.g. *Ornithodoros*) and observed to contribute to Malpighian tubules, the rectal sac and, jointly with some of the intra-yolk energids, to the anterior midgut sac and epithelium (Anderson 1973).

The cellular rearrangements of gastrulation also bring smaller, presumably proto-mesoderm cells interior and adjacent to the superficial germdisc. During this early stage of ingression these rudimentary mesodermal cells are located each side and arrayed in front of, the putative endoderm/midgut rudiment cluster (Figure 2.1.2,ivB). In some embryos I observed large, approximately equally spaced cells overlying the blastoderm and part of the germdisc, attributable to extra-embryonic ectodermal cells.

2.1.3 Germband formation and development

i) Early germband: 16-19hr AEL

Gastrulation and proliferation continues at the posterior of the germdisc, the discoidal swelling projecting inwards to a maximal 25% of the total embryo diameter (i.e. $18.75\mu\text{m}$). A broad region of high cell density that lies just in front of the gastrulation groove, an apparent growth zone, acquires more definite character as the domain of posterior-anterior differentiation (Figure 2.1.3a,i,iv). Ingressed mesodermal cells appear to migrate anterior-wards along the inner surface of the emerging germband ectoderm (Figure 2.1.3a,iB,ivB). The germband lengthens, curving around the yolk so that its boundaries prescribe a stretched kidney bean shape (Figure 2.1.3a,iii).

The cell division events captured in Figure 2.1.3a show that the scattered internal cell population is derived partly from blastomere mitosis, and part from mitosis or delamination of cells within the inner cell mass zone. Assuming that at least a fraction of the intra-yolk cells also retains the capacity to divide autonomously, multiple sources and possible fates are inferred for the ambiguous inner cell population.

ii) Mid germband stage: 20-21hr AEL

Further lengthening of the germband occurs, due to active mitosis throughout newly forming tissue. Anterior and posterior ends of the germband extend over the yolk mass to opposite sides of the dorsal

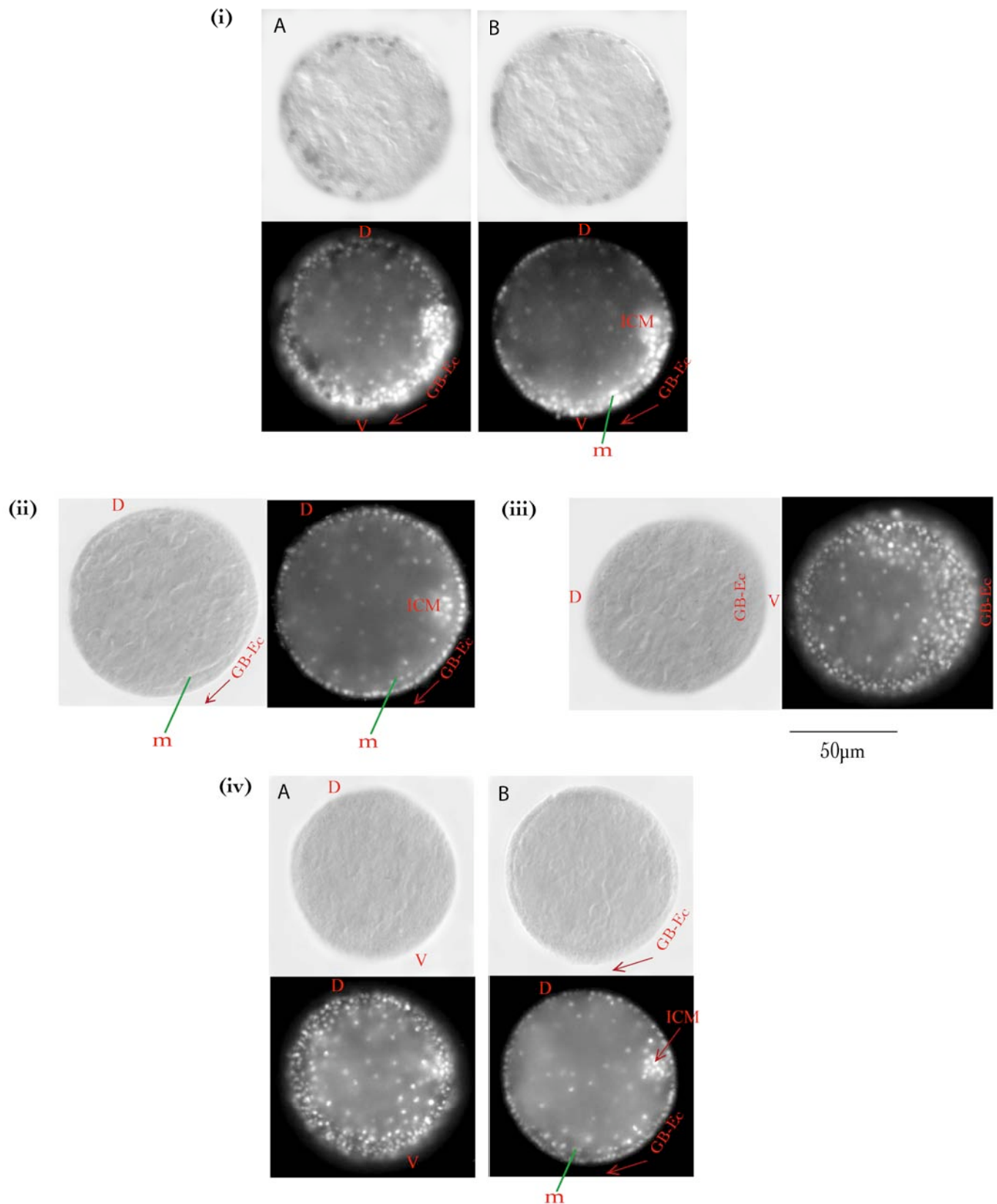


Figure 2.1.3a Early germband stage. (i) A: lateral, B: sagittal view, mesoderm ingress from posterior to anterior - pHis-III antibody. (ii) Sagittal view, distinct mesoderm layer visible in DIC. (iii) Transverse. (iv) Lateral and sagittal.

D - dorsal; V - ventral; GB-Ec - germband ectoderm; m - mesoderm; ICM - inner cell mass/endoderm.

surface, forming a horseshoe-shaped embryo when viewed in lateral cross-section (Figure 2.1.3b,ii). Attenuated extra-embryonic ectoderm fills the dorsal gap between anterior and posterior germband terminalia (Figure 2.1.3b,i-ii). An examination of sectioned embryos at multiple time points indicates that mesoderm cells continue to channel along the length of the germband, moving anteriorly as a laterally extensive, continuous layer (Figure 2.1.3b,iii). The posterior cluster of presumptive endoderm cells slightly elongates longitudinally, and is surrounded by a higher than average density of intra-yolk-type cells, possibly incorporated into subsequent endoderm-derived structures.

Between 19-20hrs AEL the first sign of segmentation appears as a subtle transverse ectodermal groove anterior to the presumptive opisthosomal rudiment. This early indentation may represent a vestigial 'disjugal furrow', an ancestral boundary in acarids that separates opisthosomal and prosomal territories. Any hypothetical disjugal furrow would not develop to be a functional body division, as major tagma are shifted in derived acariforms to give a propodosoma and hysterosoma, separated at the L2-L3 or 'sejugal' boundary, and tagma are often entirely fused in derived mites such as *Tetranychus*, in which the body is simply divided into an anterior 'gnathosoma' specialised for predation and posterior 'idiosoma' with fused carapace covering walking legs and highly reduced opisthosoma (Anderson 1973; Cloudsley-Thompson 1958; Krantz and Lindquist 1979). Alternatively, or in addition, the early furrow may delimit a posterior ambulatory limb segment (e.g. L3). Subsequently, three ectodermal grooves are apparent, more pronounced towards the posterior. Differentiation is clearly focussed in the posterior half of the germband, with growth and segment differentiation proceeding in a posterior-anterior sequence (Figure 2.1.3b,ii). This feature of morphological development is unusual among arthropods, whose segments typically appear in anterior-posterior succession from a growth zone (e.g. *Limulus*; scorpions such as *Euscorpius* and *Paruroctonus*; myriapods *Strigamia* and *Lithobius*; short-intermediate germband insects) (Anderson 1973; Farley 1998; Farley 2001; Laurie 1890; Simonnet 2005). However, diverse patterns of segmentation in the prosoma of ticks has been reported, including successive addition of more anterior segments in *Hyalomma dromadarii* (Anderson 1973).

The mesodermal cell layer continues to proliferate and extend longitudinally towards the anterior of the germband. Mesodermal cells appear to become integrated within the forming segments such that interior segmentation is visible, presumably indicating foci for somitogenesis (myogenesis and formation of other somitic mesoderm derivatives). Finally, the presumptive midgut cluster migrates to a slightly more posterior opisthosomal location, aligned with future respiratory, digestive and possibly reproductive structures (Brusca and Brusca 2003; Damen et al. 2002).

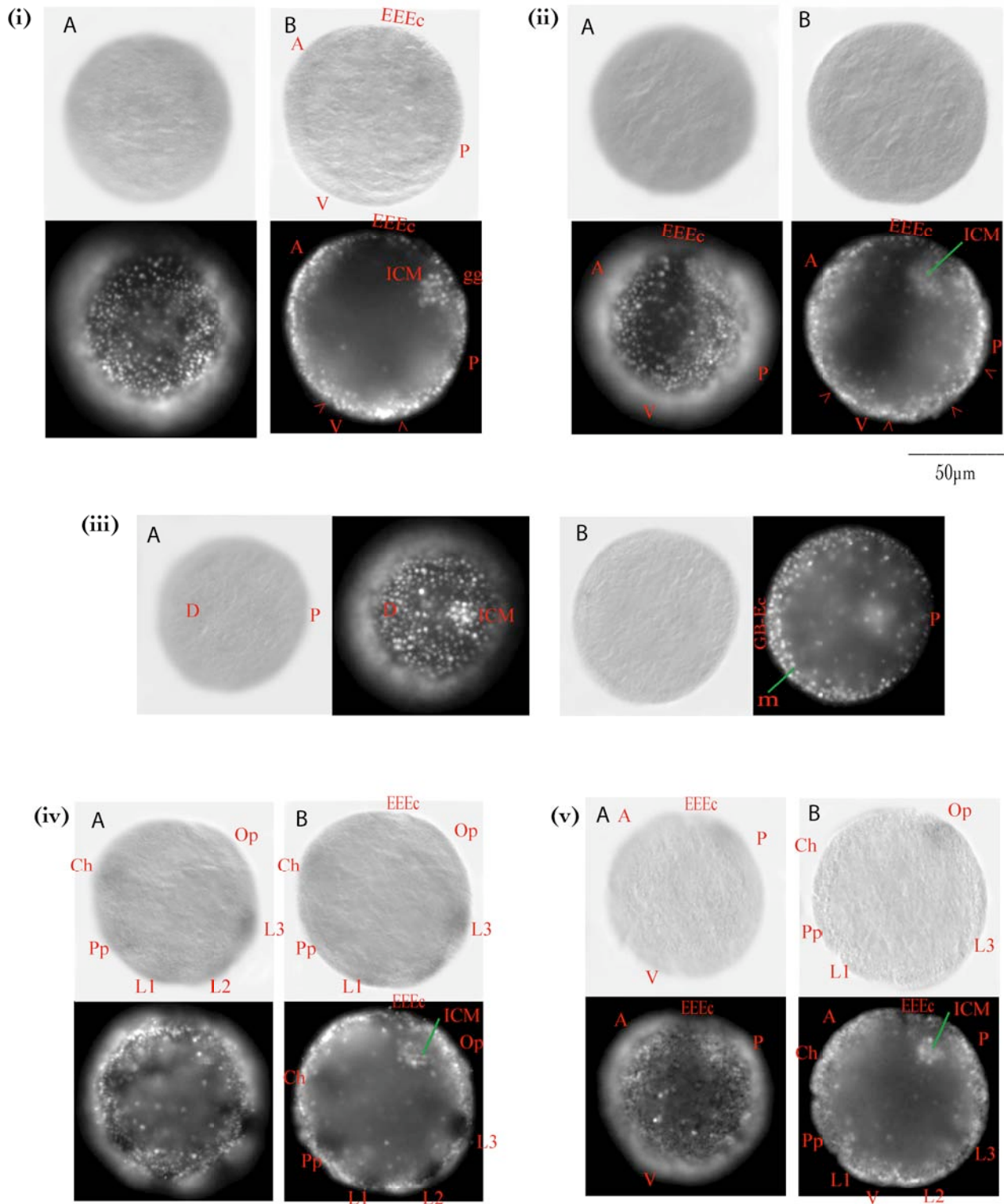


Figure 2.1.3b Mid (i - ii) and Late (iii - iv) germband stages. **Mid-germband:** (i) & (ii) Germband extends round yolk to form 'U' shape, anterior and posterior ends almost meeting on the dorsal surface. Inner cell mass migrates posteriorly. Incipient segmental groove(s) at posterior. (iii) Dorsal/posterior surface and transverse sections; mesoderm P --> A migration as broad layer. **Later germband:** (iv) & (v) Segmental grooves delimit pre-limb bud fields; anterior and posterior ends of germband meet at dorsal surface, separated by minor extra-embryonic ectoderm. (iv) Faint *Dll* mRNA transcription, highlighting earliest limb primordia of L3 - Ch segments. (v) Faint *Ubx* transcription, labelling the opisthosomal growth zone region, overlying the inner putative endoderm cell cluster.

D - dorsal; V - ventral; A - anterior; P - posterior; EEEc - extra-embryonic ectoderm; m - mesoderm; ICM - inner/endoderm cells; Ch - cheliceral primordium; Pp - pedipalp; L - walking limb; Op - opisthosomal ectoderm; > incipient groove site.

iii) Late germband: 21-22hrs AEL

In later germband development the canonical embryonic complement of 5 prosomal limb segments are specified; namely from the anterior, chelicerae, pedipalps and three pairs of walking limbs - the final fourth pair of legs not developing externally until after the first larval moult (Figure 2.1.3b,iv-v). This 5-segment stage is preceded very transiently by a phase apparently lacking the most anterior, cheliceral segment. Delayed cheliceral specification is a synapomorphy of the Chelicerata, but the delay is not very clear in *Tetranychus*, especially as the prosoma appears to segment in rapid posterior (walking limb 3) to anterior (pedipalp) succession.

The pre-cheliceral, or cephalic, lobe becomes more distinct and the germband extends to its maximum length such that the extra-embryonic ectoderm occupies a minimal gap between the dorsal, adjacent posterior and anterior ends (Figure 2.1.3b,ivB). The putative endoderm cell cluster completes its short migration to beneath the posterior opisthosoma and behind the main growth zone (Figure 2.1.3b,iv-v). This is close to the site of a ventral midline invagination that arises independent of the transient gastrulation groove to form the proctodeum. The stomodeum probably also invaginates independently within the pre-cheliceral lobe ectoderm midline, posterior to a region of ocular/labral ectoderm.

2.1.4 Appendage development

Once the germband is completely segmented and pre-limb bud domains delineated by fine ectodermal grooves, a dynamic sequence of events unfolds, dominated by limb growth and morphogenesis, ventral closure, germband contraction and dorsal closure.

i) Early limb buds: 23-24hrs AEL

Early limb buds (Figure 2.1.4a,i-iii) are flattened structures, resembling rounded paddles in outline and overlapping slightly at adjacent posterior borders due to cell division and growth relative to the previous late germband stage. The Pp to L3 limbs are directed posteriorly and the Pp-L2 limbs are 2-3µm longer than the L3 or Ch appendage, of which this last is the broader by a few µm. A ventral sulcus ~20µm wide and tapering posteriorly separates the two sides of the bilateral germband (Figure 2.1.4a,i).

A small antero-lateral lobe is to be seen on each pedipalp, projecting at ~50° from the main palp axis (Figure 2.1.4a,i), as similarly observed in the oribatid *Archegozetes longesitosus* (Thomas and Telford

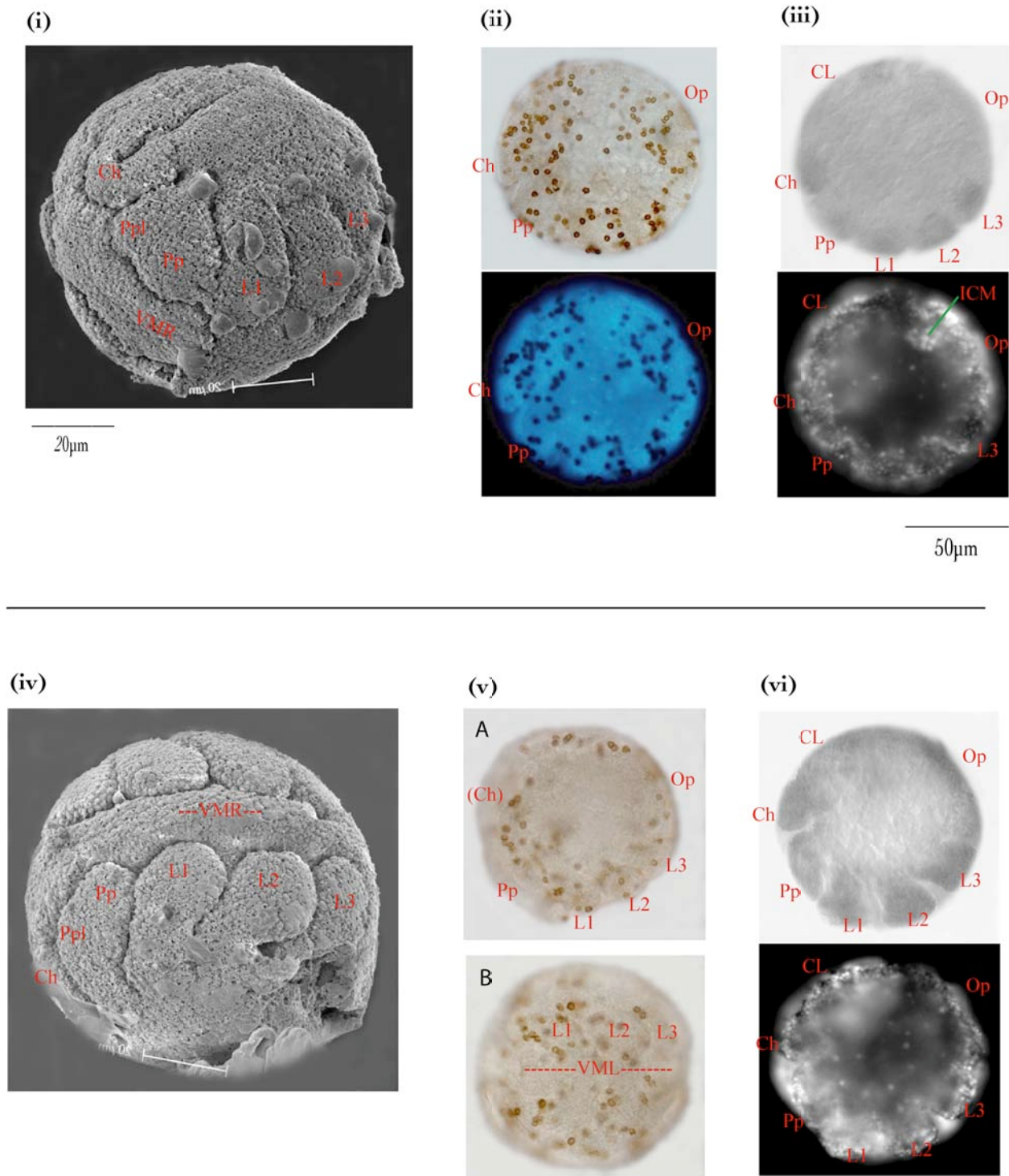


Figure 2.1.4a Early (i - iii) and Late (iv - vi) limb bud stage embryos.

Early limb buds: (i) SEM; (ii) pHIS-III antibody labelling of nuclei - DIC and DAPI; (iii) Dll mRNA transcription pattern, marking limb buds. Late limb buds: (iv) SEM; (v) pHIS-III antibody labelling nuclei within germband and elongation limbs, from lateral (A) and ventral (B) views; (vi) Dll and Ubx mRNA expression in limbs and opisthosoma respectively. Anterior to left.

Ch - chelicera; Pp(l) - pedipalp (lobe); L1 - pair 1 walking limb; Op - opisthosoma; VMR - ventral midline ridge; VML - ventral mid-line; CL - cephalic lobe; ICM - inner cell mass.

1999). In sagittal section (Figure 2.1.4a,iii) it is clear that the germband still curves almost completely around the yolk, the anterior and posterior ends approaching one another on the dorsal surface.

ii) Late limb buds: 25-30hrs AEL

The limb buds become more elongate, with the proximo-distal axis broken into a proximal ventrally oriented section and a posteriorly directed distal region. The limbs remain flattened, aligned with curvature of the germband, and are closely opposed to one another along both anterior and posterior margins. The ventral gap is reduced to ~10-12µm due to limb extension and active ventral closure, the process of ventral contraction creating a medial ridge, most pronounced between limbs L1-L3 (Figure 2.1.4a,iv). The pedipalps and chelicerae remain broader and more rounded than the walking limbs, which are narrower than in the early limb bud stage. Continued somatic growth is evidenced by mitotic division throughout the germband, shown *via* antibody staining for phospho-Histone-III at dividing centromeres (Figure 2.1.4a,v).

iii) Ventral ridge closure: 30-39hrs AEL

During contraction of the ventral ridge, appendage morphogenesis forms more strongly posterior-directed limbs that are longer, narrower, more substantial in lateral cross-section and have rounded distal podia (Figure 2.1.4a,i-iv). The ventral gap closes such that only the ventral ridge separates limbs of the right and left sides, at which point the ridge begins to contract further and disappear. During this later phase, the first sign of L4 appendage rudiments are just detectable on the L4 segment as faint, curved grooves encompassing slight ectodermal swellings relative to the adjacent limbless opisthosomal ectoderm (Figure 2.1.4b,ii).

In the future gnathosomal area, the pedipalp is significantly elongated and its lobe considerably enlarged and directed horizontally towards the ventral midline (Figure 2.1.4b,iii-iv). Chelicerae are elongated, slightly bilobed distally and approach one another closely, with an anterior-ward rotation as the ventral ridge is eliminated (Figure 2.1.4b,iv). The stomodeum is apparent at this stage at the ventral midline, just below the chelicerae and approximately in line with the Pp lobes, having originated at an ingression site independent of the transient gastrulation groove (Anderson 1973). The labrum is also apparent, albeit quite faintly, at the ventral midline of the pre-cheliceral lobe, inserted just above the chelicerae (Figure 2.1.4b,iiiA,ivA).

The posterior cell mass is less clearly defined beneath the opisthosoma at this stage, presumably due to dynamic incorporation within the developing gut and gonads, and it seems that as the ventral ridge fades, the terminal opisthosomal and cephalic lobe boundaries move apart at the dorsal surface, indicating incipient or preliminary longitudinal germband contraction (Figure 2.1.3b,iiiA).

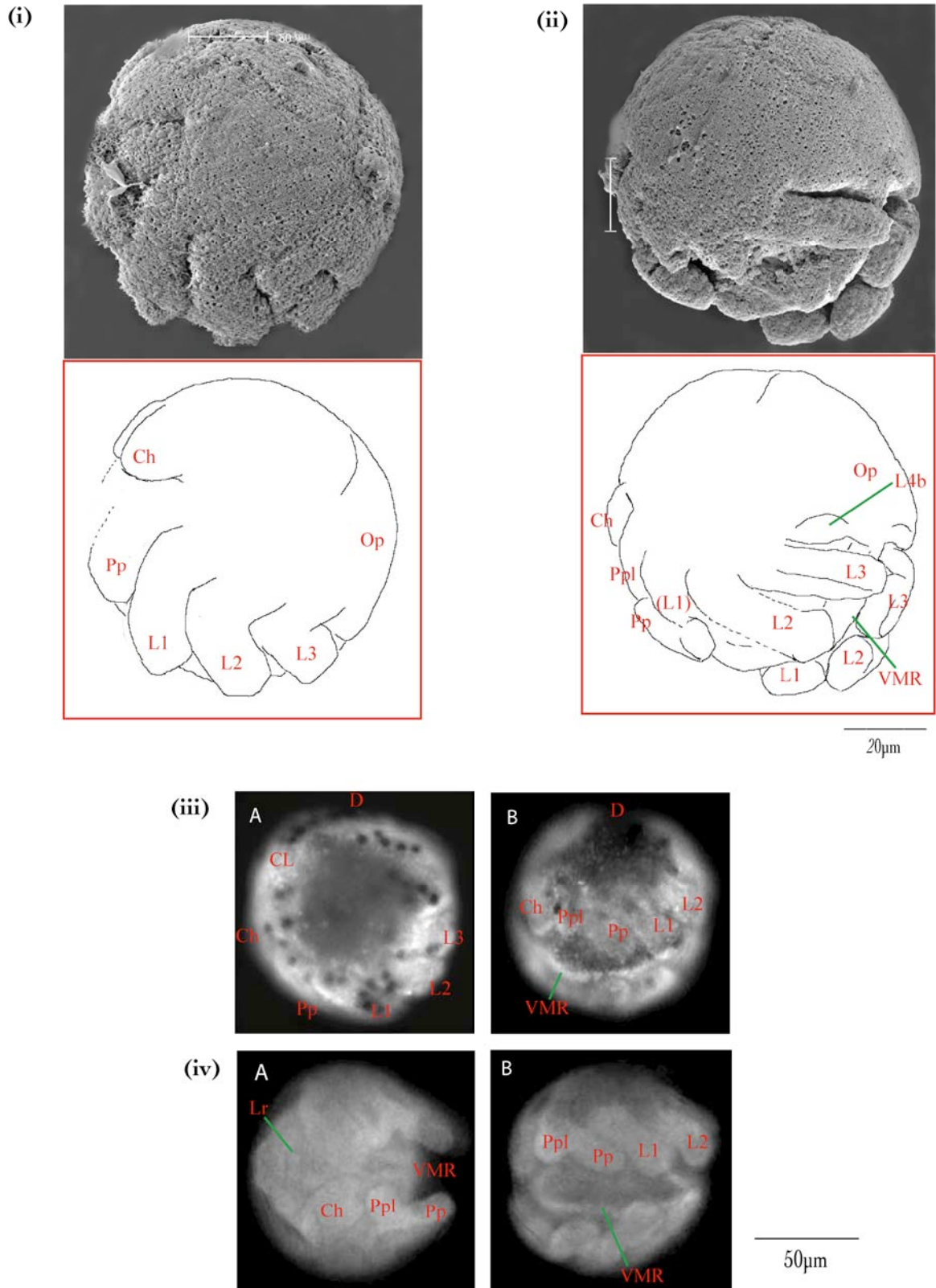


Figure 2.1.4b Ventral ridge closure (40 - 45hrs AEL). (i) SEM lateral view; (ii) SEM postero-lateral view, with line drawings. (iii) DAPI - pHis-III labelling; lateral (A) and ventro-lateral (B) views. (iv) Anterior (A) and ventral (B) views, DAPI stained embryos. Labrum apparent in (ivA). Anterior left. Ch - chelicera; Pp(l) - pedipalp (lobe); L1 - pair 1 walking limb; Op - opisthosoma; VMR - ventral midline ridge; VML - ventral mid-line; CL - cephalic lobe; ICM - inner cell mass; L4b - pair 4 limb bud stump.

iv) Ventral closure and limb elongation: 40-45hrs AEL

Elimination of the ventral ridge is followed by complete ventral closure and extensive limb growth such that members of each pair interlock in very close apposition (Figure 2.1.4c,i-iiD). The L1-L3 appendages exhibit two or three shallow annulations along their proximo-distal axes, resembling early traces of limb segmentation. The distal part of the walking limbs and pedipalp remain enlarged relative to the narrower limb axis, and in addition to the cheliceres remaining distally bilobed as previously, the Pp-L3 distal appendages also become distinctly bilobed (Figure 2.1.4c,i-ii,iiiC-D). These bi-partite distal segments presumably reflect underlying morphogenetic and organogenetic events related to the sensory complexes that develop at the extremity of each limb (e.g. empodia, section 2.2)(Flechtman 2002; Zhang, 2000; Brussels, 2006; Grandjean, 1935; André, 2003; Thomas, 1999). The labrum is prominent at the midline of the pre-cheliceral lobe, and the chelicerae and pedipalp lobes are rotated near-horizontally to interlock at the midline and obscure the stomodeum directly beneath (Figure 2.1.4c,iii). Of note, the labrum is distally bilobed, suggesting appendage affinity and supporting gene expression studies in arthropods and functional studies in *Tribolium castaneum* that propose a segmental appendage origin for this long-debated structure (Popadic et al. 1998; Haas et al. 2001a; Haas et al. 2001b) (Thomas and Telford 1999). The L4 rudimentary limb bud is only evident as an oval swelling tucked behind the L3 limb (Figure 2.1.4c,ii,iiiE), its proximo-distal development suppressed as though held in a paedomorphic state (de Beer 1940).

The germband begins to contract longitudinally during this stage, shown by minor compression of limbs as they are accommodated within a shorter antero-posterior axis, and by a widening dorsal domain of extra-embryonic ectoderm, formed as the anterior and posterior ends of the germband retreat away ventrally (Figure 2.1.4c,iiiE-F). The opisthosoma tapers away from its junction with the prosoma and is composed of segments reduced in both size and number compared to other Arachnids, preceding a terminal telson (Anderson 1973; Thomas and Telford 1999). Cells between the proposed 2nd opisthosomal (Op2) segment and terminal area are at lower density, and no transverse segmentation is discernable. This may be because the opisthosoma subsequently contracts, and even if the five opisthosomal segments typical to acarids were present at any molecular or faint morphological level, they would be obliterated during the secondary reduction of this body region in *Tetranychus*.

v) Germband contraction and dorsal closure: 45-50hrs AEL

The germband is fully contracted to the ventral side during the latest stages of embryonic development. The Pp-L3 limbs elongate further and are traversed at multiple points along their

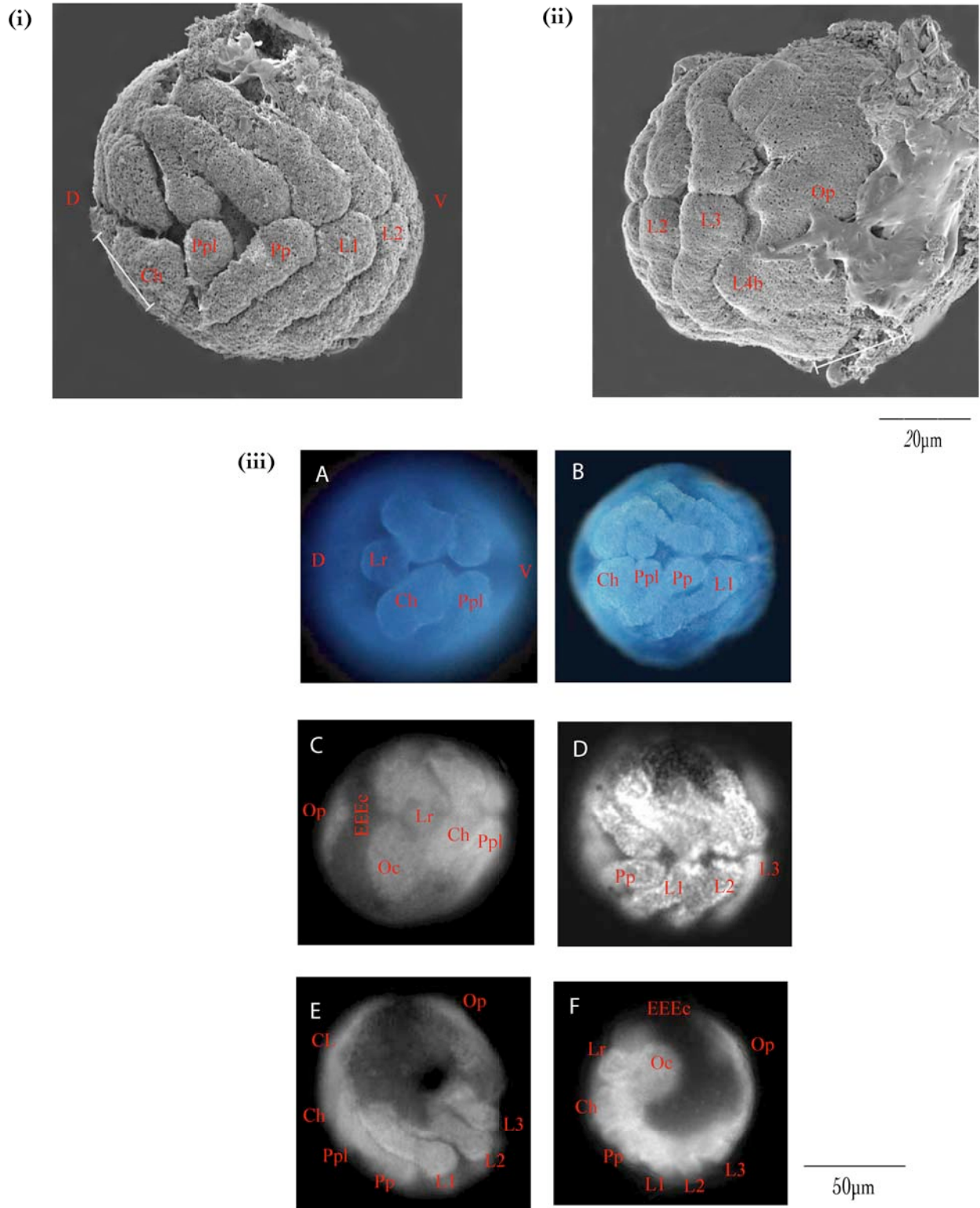


Figure 2.1.4c Ventral closure: 40 - 45hrs AEL. (i) SEM ventral; (ii) SEM posterior. (iii) DAPI stained embryos: A - bilobed labrum and chelicerae; B - antero-ventral view; C - cephalic lobes, labrum, chel'ae; D - ventro-lateral; E - lateral; F - sagittal, cephalic lobe and labrum prominent. Anterior left.
 Ch - chelicera; Pp(l) - pedipalp (lobe); L1 - pair 1 walking limb; Op - opisthosoma; VMR - ventral midline ridge; VML - ventral mid-line; CL - cephalic lobe; L4b - pair 4 limb bud stump; Oc - ocular lobe; EEEc - dorsal extra-embryonic ectoderm; D - dorsal.

proximo-distal axes by annulations signifying joints between successive articulating segments; proximal trochanter to distal tarsus (Figure 2.1.4d,i-ii) (Flechtman and Knihinicki 2002; Jeppson et al. 1975; Brussels 2006). Mitotic activity within jointed L1-L3 limbs supports an embryonic growth mechanism based upon increased somatic cell number, rather than a mechanism dominated by increased cell size, as proposed by Anderson (1973) for basal acarids (Anderson 1973). The L4 limb rudiment (Figure 2.1.4c,i) attains a more pronounced and extended form, shorter but similar in shape to the early Ch-L3 limb buds in being curved towards the posterior (Figure 2.1.4a,i).

In the gnathosomal region the labrum is reduced, presumably due to apoptosis or cellular rearrangement. The chelicerae continue to migrate, arriving at a supra-stomodeal (pre-oral) position in which the appendages are juxtaposed and can fuse basally as formation of the stylophore and its internal apparatus proceeds. Pedipalp lobes, having extended and come into close apposition, also begin to fuse as they form the infracapitulum, onto the ventral surface of which the stomodeum then migrates.

Dorsal closure is initiated once the germband is fully retracted to the ventral surface, opisthosomal and prosomal body tissue moving dorsally over the inner yolk mass to enclose it (Figure 2.1.4d,iii). The coxal bases of Pp-L3 limbs move to a higher dorso-lateral position, the closure thus also providing an indirect mechanism for accommodating further limb growth and elongation. Completion of dorsal closure reveals a generalised body plan typical of the Acariformes: the chelicerae and pedipalps comprise the 'gnathosoma' - the anterior division of the spider mite body dedicated to feeding and sensory functions, and the remainder of the body is termed the 'idiosoma', comprising walking limbs and an extremely reduced, fused opisthosoma.

2.2 Post-Embryonic Development

2.2.1 Prelarva

i) Early prelarval features: >50hrs AEL

Prior to hatching a prelarval stage exists, its development proceeding entirely within the chorion and vitelline membranes. The appendages are very closely juxtaposed, and a near-straight linear junction between left and right appendage pairs is a defining characteristic of this stage (Thomas and Telford

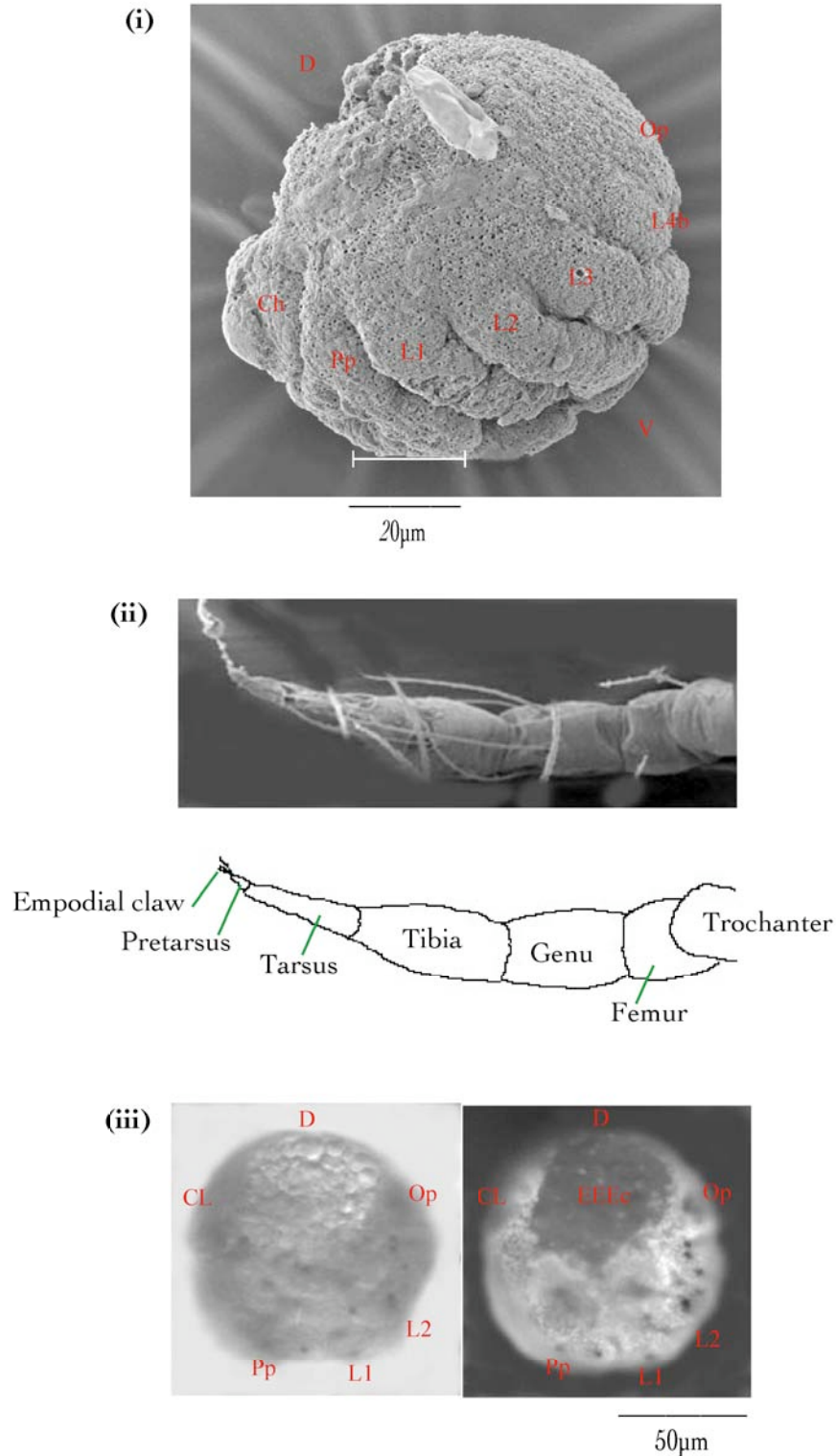


Figure 2.1.4d Germband contraction and dorsal closure: 45 - 50hrs AEL. (i) SEM lateral; (ii) mature walking limb segments (iii) DIC and DAPI, dorso-lateral view of dorsal closure. Anterior left. Limb segment annotation in (ii) is based on Jeppson (1975).

Ch - chelicera; Pp(l) - pedipalp (lobe); L1 - pair 1 walking limb; Op - opisthosoma; VMR - ventral midline ridge; VML - ventral mid-line; CL - cephalic lobe; L4b - pair 4 limb bud stump; Oc - ocular lobe; EEEc - dorsal extra-embryonic ectoderm; D - dorsal.

1999). Superficial surface features are quite smooth and indistinct, as tissue formed by continued growth and morphogenesis presumably generates positive pressure against, and hence uniform opposing resistant force in response from, the egg membranes. Coxal bases of the palps and walking appendages occupy high dorso-lateral positions, having moved as the prosomal wall extended and dorso-ventral musculature was rearranged at the end of dorsal closure. Joints are still visible as annuli distributed along proximo-distal limb axes, though only faintly due to compression from external membranes. By this stage there are no clear morphological signs of the L4 appendages, although limb rudiments of some form presumably remain, and the labrum has been completely re-incorporated into other cephalic tissues.

Acarid chelicerae are typically paired pincers or styliform appendages, with fused or moveable joints, adapted for predation of various kinds. Spider mite chelicerae are highly modified, being adapted specifically to a phytophagous lifestyle, and the chelicerae of *Tetranychus urticae* are no exception (Boudreaux 1963). The cheliceral bases of *Tetranychus* fuse to form a bulbous, semi-retractable stylophore, the outline of which is visible in the prelarva with a medial margin suggesting that fusion of the previously separate, bilateral chelicerae is not yet complete (Figure 2.2.1,i). Fusion of the pedipalp lobes (coxal processes) forms a beak-like lower part of the gnathosoma, termed the infracapitulum. Its surface outline is clear, with a vestige of previous segmentation due to incomplete fusion of adjacent lobes. The pedipalp bases (coxa and trochanter) are located lateral to the palpal lobes, either side of the infracapitulum (Figure 2.2.1,i).

ii) Late prelarval features: >60hrs AEL

Detailed sensory features of the gnathosoma, both chelicera- and palp-derived, attain clearer morphological definition in the late prelarva. As well as sensory organogenesis and maturation of gnathosomal structures this ‘pre-hatching’ stage is characterised by final appendage elongation and dark red pigmentation of the simple, bilateral eyes, making them visible through the chorion (Figure 2.2.1, ii-iii).

Sensory structures primarily take the form of setae on the dorsal integument and both dorsally and distally on the limbs (Brussels 2006; Grandjean 1948; Zhang and Jacobson 2000). On the palps, both chemo- and mechano-sensory structures become apparent, as does a tarsal silk-producing organ, adapted by modification of a specialised hollow seta (empodium) (Krantz, 1979; Návia, 2004). Within the stylophore, a pair of fixed digits are present as thin medio-lateral, posteriorly hooked projections (peritremes), accompanied by a pair of very elongated, proximally recurved ‘stylet’ chelicerae that can be protracted and interlock as an organ for piercing, sucking and transfer of saliva via a duct connected to salivary glands in the idiosoma. The infracapitulum has apical mouth lobes,

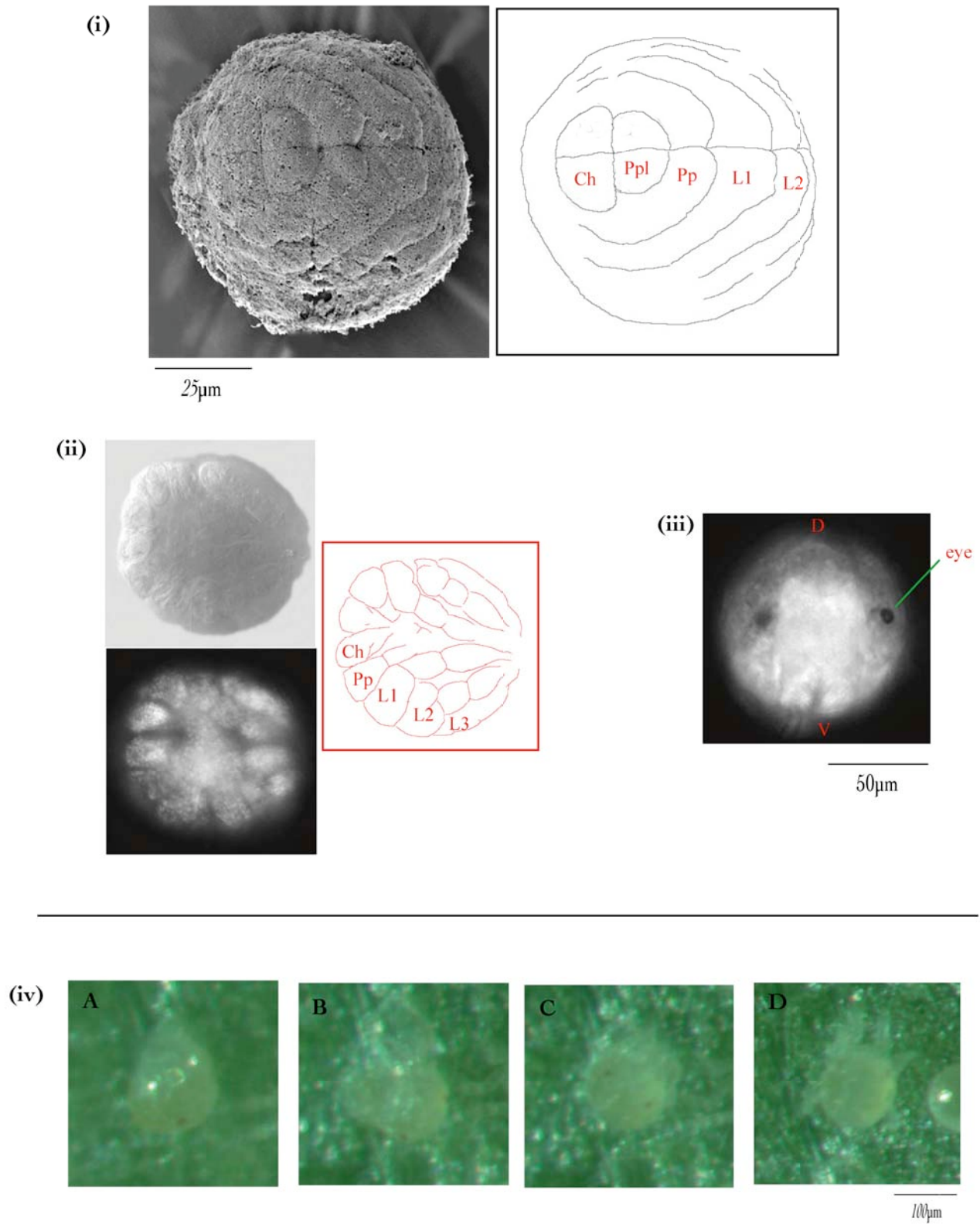


Figure 2.2.1 Prelarva (i - iii) and hatching sequence (iv). (i) SEM and parallel drawing, antero-ventral view. (ii) DIC and DAPI, ventral view of appendages with joints and chaetae. (iii) DAPI anterior view, showing eye pigmentation prior to hatching. (iv) A - D temporal sequence of one hexapod larval hatch.

Ch - chelicera; Pp(I) - pedipalp (lobe); L1 - pair 1 walking limb; D - dorsal; V - ventral.

used during feeding to form a seal around the protracted stylets, and a small oral opening on its ventral surface, leading to the pharynx and oesophageal canal (Crooker et al. 2003).

2.2.2 Hexapod larva: >70hrs AEL

Tetranychus urticae larvae hatch as pale yellow-green, translucent mites with prominent red eyes and three pairs of legs; the fourth pair of walking limbs is not yet formed (Figure 2.2.2,i-ii). Hatching occurs after 70hrs at 25°C, emergence taking a few minutes as hexapod larvae push egg material away with the infracapitulum apex and anterior limbs (Figure 2.2.1,iv). Emergence is followed by a feeding phase in which larvae acquire a darker green colouration, especially in the posterior idiosoma, and food particles and metabolites become visible through the translucent body wall as two persistent dorso-lateral patches, after which the 'two-spotted' spider mite is named (Bynum and Porter 2006; Collyer 1998).

The larval carapace is ~90µm long, very rounded and marked by integumentary striae primarily arranged transversely, with bilobed striations traversing the anterior-most idiosoma (Figure 2.2.2,i-iii). The orientation of striation is not entirely transverse (see Figure 2.2.2,iiiA), and indeed, the diagnostic character for the genus *Tetranychus s. str.* is longitudinal striae forming a diamond shape between two posterior-central pairs of setae (Flechtmann and Knihinicki 2002). The larval stylophore (fused cheliceral bases) seems reduced in length relative to that of the adult, larval palps being of almost equal length to the infracapitulum (Figure 2.2.2,iiiB). Boudreaux (1968) suggests that a complete complement of somatic cells are formed by the larval instar, and size increase in subsequent nymphal and adult stages result from cell size increase rather than somatic mitosis (Boudreaux 1963). This theory is supported in this study by a lack of evidence for mitotic activity in the few cases when larvae or nymphs were stained with a phospho-Histone-III antibody that marks dividing centromeres.

2.2.3 Nymphal stages

Tetranychid mites have two post-larval instars, the protonymph and deutonymph (Figure 2.2.3). The final eight-legged body plan typical of chelicerates is reached in the protonymph stage, and retained throughout further ontogeny (Anderson 1973). Lack of a third, tritonymph stage represents a derived, paedomorphic state relative to most acarids, whose life cycle includes three post-larval instars prior to adulthood (Florida 2005; Krantz and Lindquist 1979).

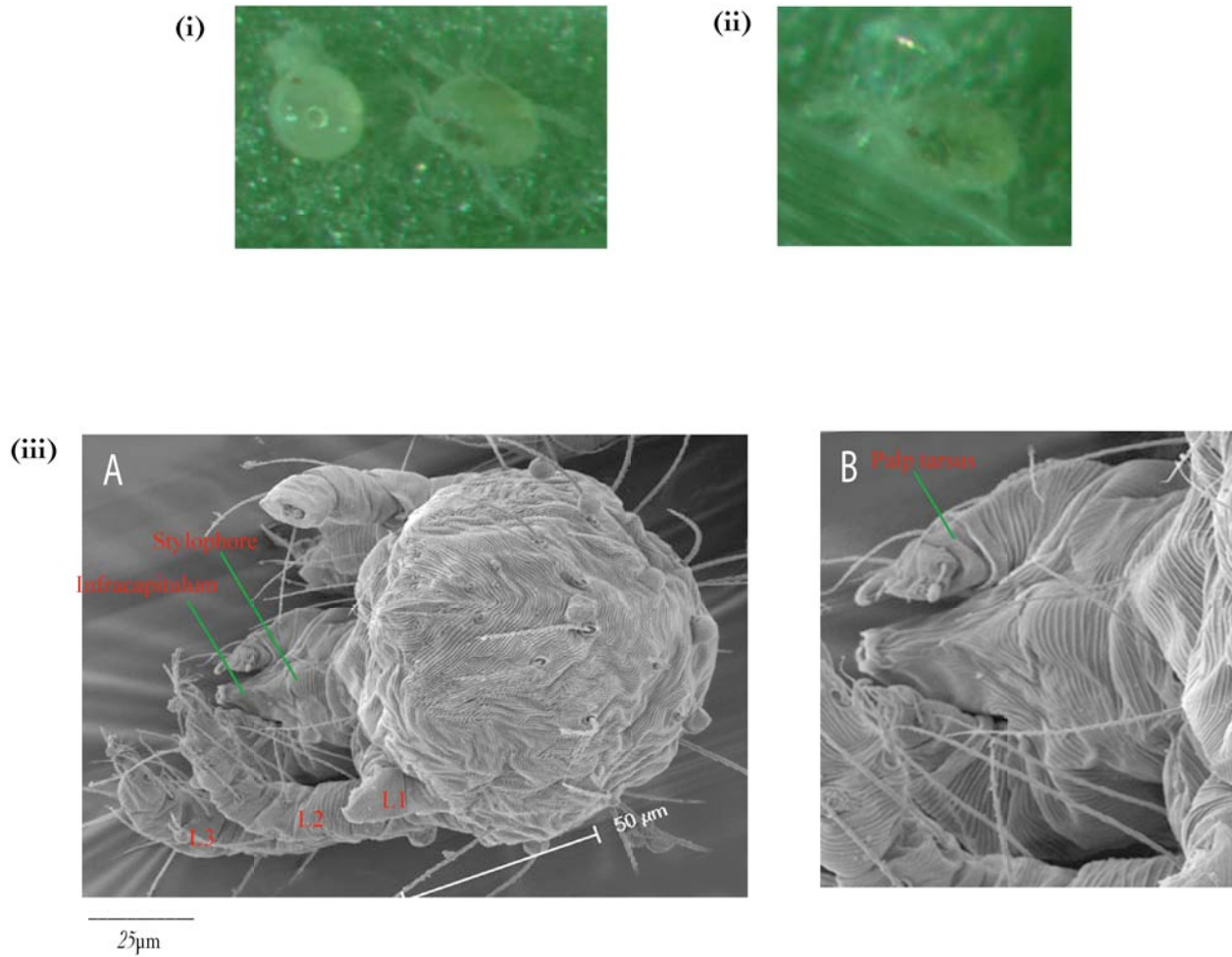


Figure 2.2.2 (i) Hexapod hatchling, postero-dorsal view. (ii) Hexapod, lateral view. (iii) SEM: A - dorsal chaetotaxy; B - enlargement of gnathal region.



Figure 2.2.3 Spider mite nymphal stages. (i) Pronymph (150-175μm) and deutonymph (250-300μm). (ii) Female (400-450μm) with pronymph, and (vi) female with deutonymph, for size comparison. Images from Collyer [www \(1998\)](http://www.collyer.com); U. of Florida [www \(2005\)](http://www.ufl.edu).

The hexapod - protonymph transition involves anchoring of the larva to a leaf or web, inward folding of the walking limbs, and passage through a static 'protochrysalis' stage before ecdysis as the first nymphal instar. The new cuticle and fourth pair of walking legs form within the original larval cuticle and limbs, this and segment addition being indicative of anamorphic development (Boudreaux 1963; Collyer 1998; Minnesota 2006). After a feeding period lasting up to 2 days, the protonymph transforms into a 'deutochrysalis', which moults to give the deutonymph instar (Boudreaux 1963; Bynum and Porter 2006; Collyer 1998; Minnesota 2006). The deutonymph is similar to the protonymph, being eight-legged and oval in form, but is larger and may bear extra pairs of tactile setae on the postero-dorsal idiosoma (Zhang and Jacobson 2000). Protonymph body length is 150-175µm, and deutonymphs measure from 250-300µm; still at least 100µm smaller than an adult female (Figure 2.2.3). Deutonymphs are slightly longer-lived than protonymphs, typically living 72 hours relative to 48 hours for protonymphs (Collyer 1998; Minnesota 2006).

2.2.4 Adult stage

Reproductive spider mite adulthood is reached *via* a teliochrysalis moult, 5-15 days after egg-laying, the time to maturity dependent on environmental conditions. Development is accelerated between 25-37°C (77-98°F) and in low humidity (Boudreaux 1963; Bynum and Porter 2006). Females and males are sexually dimorphic, most clearly in their overall body shape and genital structures, but also in colouration and subtle aspects of sensory setal number, arrangement and form (termed 'chaetotaxy'; further details in Appendix 1) (Návia 2004; Flechtmann 2002). Within Acariformes, all types of setae are made of birefringent actinochitin, hence the alternative suborder name, 'Actinotrichida' (Krantz and Lindquist 1979).

i) Female and male colouration and body form

Adult females are generally green or green-yellow and somewhat translucent, with two dark bilateral spots comprised of ingested material and metabolic waste accumulations (Figure 2.2.4a,i). Females may take on a brown or orange-red colour (Figure 2.2.4a,ii), particularly in response to decreased photoperiod, nutrient availability or temperature in the autumn. Such environmental conditions induce body fat deposition in preparation for diapause, causing green 'two-spotted' colouration to be replaced with brown. Diapausing adult females hibernate in temperate climates from late autumn to early spring, in leaf litter, ground litter or bark. Males are not known to diapause, but maintain a green or yellow-green colour throughout their life, also acquiring diagnostic bilateral spots as food particles accumulate beneath the translucent exoskeleton (Figure 2.2.4a,iv). Adult body size is at

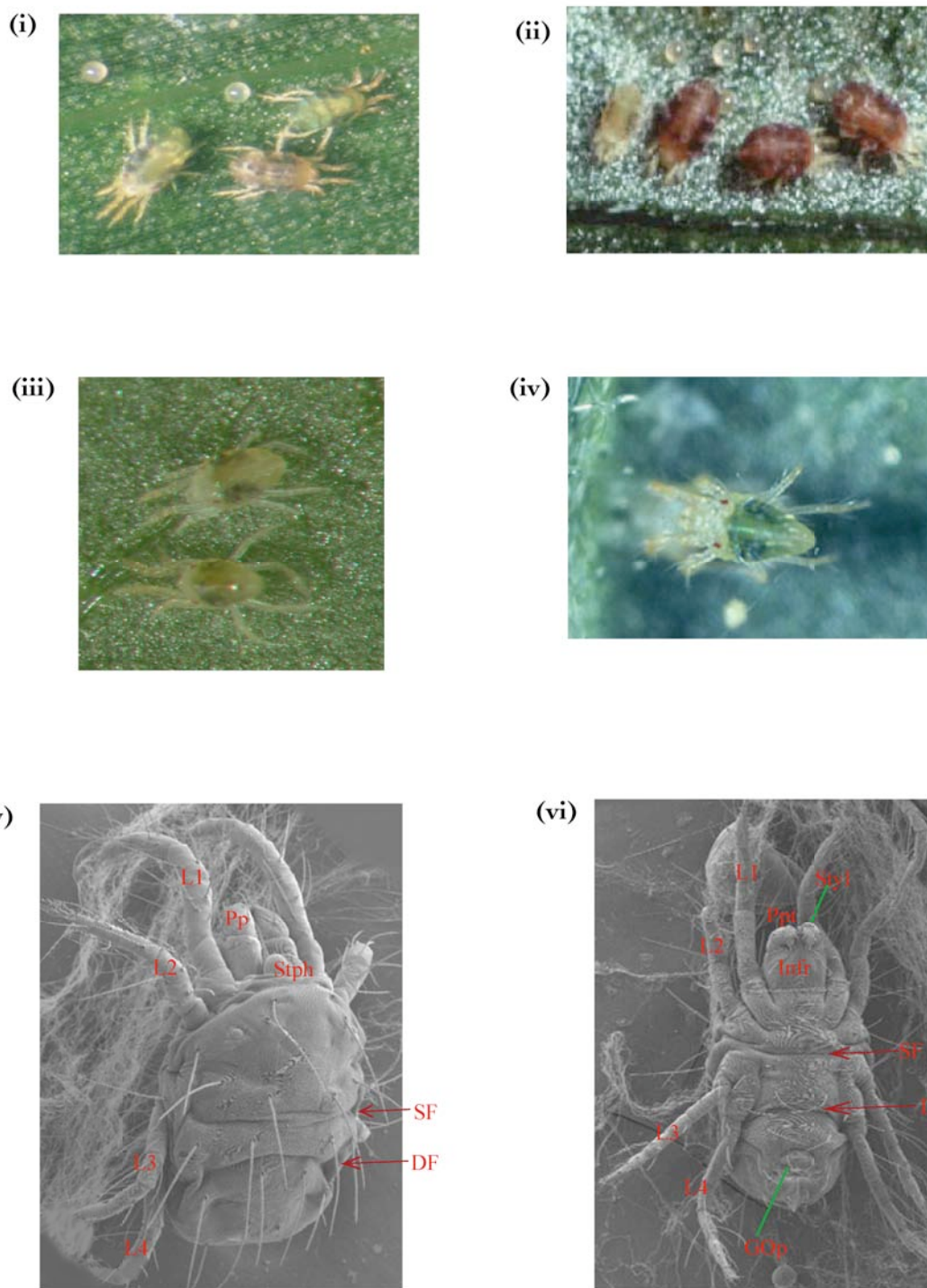


Figure 2.2.4a Adult spider mite morphology and chaetotaxy. (i) Early season females; (ii) Late season females, brown colouration. (iii) Size comparison of female (above) and male (below). (iv) Male, with pointed abdomen. (v) SEM - female dorsal chaetotaxy and detailed morphology; (vi) SEM - female detailed ventral morphology. Disjugal (Pro-Op) and sejugal (L2-L3) furrows visible.

Photos from Collyer [www \(1998\)](http://www.collyer.ca); Agriculture Canada; SEM from Bielza (2005).

Ch - chelicera; Pp(t) - pedipalp (tarsus); L1 - pair 1 walking limb; Stp - stylophore; Styl - cheliceral stylets; Infr - infracapitulum; GOp - genital operculum; SF - sejugal furrow; DF - disjugal furrow

least 50-100µm greater than the preceding deutonymph stage, dependent on sex: female body length reaches 400-450µm, whereas males are smaller, being 300-350µm long (Figure 2.2.4a,iii). Both sexes have relatively longer legs than nymphs and elliptical bodies in lateral view, but the female carapace is oval-shaped, distinguishing them from the male, which is tapered towards the caudal, posterior end (Figure 2.2.4a,iii-iv).

ii) Adult limb and body segmentation

Spider mite walking limbs diverge from a typical arthropod segmentation pattern in having an extra segment (genu) between the femur and tibia (Jeppson et al. 1975). The ventral undersurface (Figure 2.2.4a,vi) reveals vestiges of segmentation in the form of a disjugal furrow at the presumptive prosomal-opisthosomal boundary (overtly visible dorsally in the Opilioacariformes) and a sejugal furrow between the L2 and L3 segments, the latter being a derived character of the Acariformes (Anderson 1973; Anderson 1999). The prosomal carapace (Figure 2.2.4a,v) seems to lack overt segmentation, the body being divided according to functional appendage type into an anterior gnathosoma (chelicera/palpal lobe-derived mouthparts and palps) and posterior idiosoma (walking limbs) (Brusca and Brusca 2003; Crooker et al. 2003). However, scanning electron microscopy of the carapace has confirmed traces of both a disjugal (prosomal-opisthosomal) and sejugal (L2-L3) boundary furrow (Bielza 2005; Brusca and Brusca 1990). The prosomal-opisthosomal fusion boundary transverses the whole dorsal surface but the sejugal furrow is less marked and fades medially (Figure 2.2.4a,v). As previously mentioned, integument striations are largely transverse, but lobed behind the stylophore and form a diagnostic diamond-shape between specific posterior setal pairs (Flechtmann and Knihinicki 2002).

iii) Gnathosoma functional morphology: Phytophagy

The adult gnathosoma is identical in general structure to the larval, except for increased size and approximately 30% increased palp length relative to the infracapitulum (Figure 2.2.4b,iii-iv). Regarding palp tarsal chaetotaxy, a broad terminal sensory structure termed a 'eupathidium' is present, secondarily adapted to function as a spinnaret for silk production. Two smaller lateral eupathidia and dorso-apical sensory setae termed 'solenidia' are also located on the distal tarsus, serving as gustatory and olfactory receptors respectively (Brussels 2006; Flechtmann and Knihinicki 2002; Krantz and Lindquist 1979; Zhang and Jacobson 2000). Combined with tactile setae dedicated to providing mechanosensory input, the palps are well equipped for locating appropriate food sources and spinning silk webs.

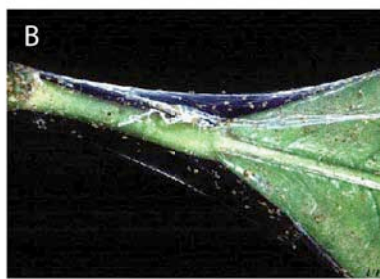
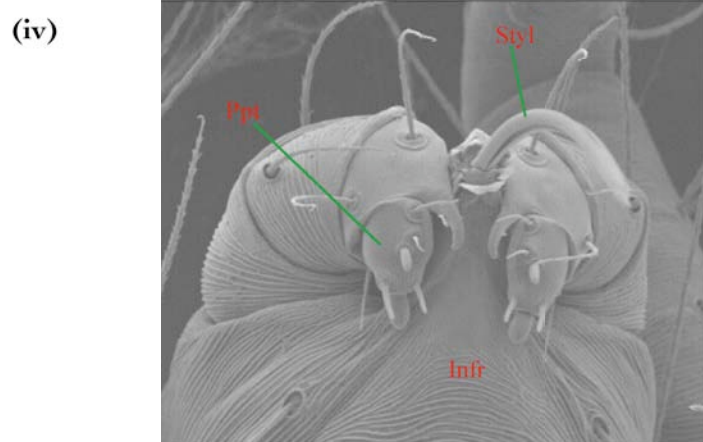
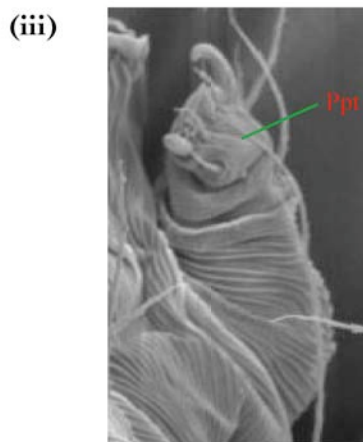
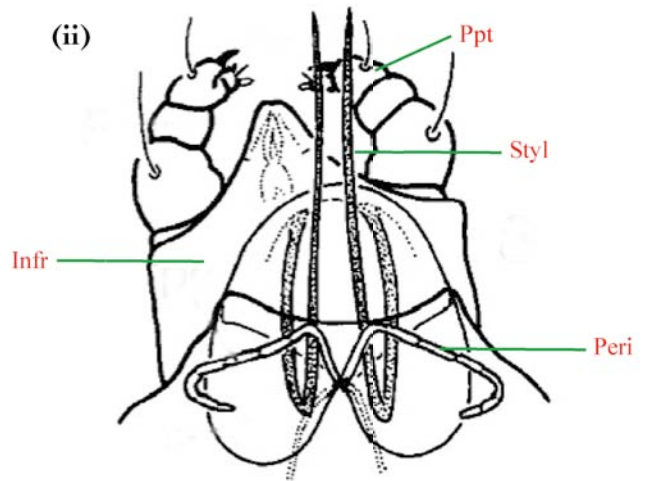
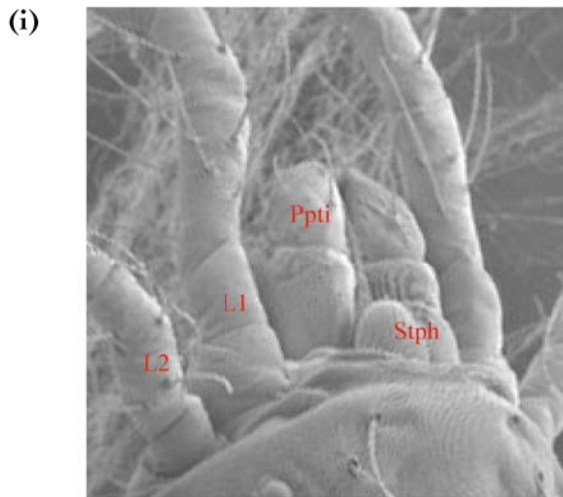


Figure 2.2.4b Spider mite gnathosoma, phytophagy and biological control. (i) SEM dorsal view gnathosoma, with (ii) line drawing to show major structures. Hexapod (iii) and Adult (iv) palp tarsi, infracapitulum and in (iv), chelicerae. SEM images (i) and (iv) from Bielza (2005).

(v) Plant damage due to spider mite feeding: A - leaf mesophyll necrosis, B - late infestation webs.

(vi) A biological control species, the predatory acarid *Phytoseiulus persimilis*. Photos: Agriculture Canada; Collyer (1998).

Ch - chelicera; Ppt/ti - pedipalp tarsus/tibia; L1 - pair 1 walking limb; Stph - stylophore; Styl - cheliceral stylets; Infr - infracapitulum; Peri - peritreme.

Cheliceral stylets (partly everted from the stylophore in Figure 2.2.4b,iv) interlock to form a unitary, tube-like process, supported by the apex of the infracapitulum. Protractor muscles evert the stylets and provide force required to pierce plant tissues. Salivary glands in the idiosoma are linked to the buccal cavity via salivary ducts, and the stylet tube may convey saliva to facilitate ingestion and primary digestion at puncture sites (Krantz and Lindquist 1979). Spider mites may penetrate 18-22 leaf cells per minute when feeding, sucking out cell contents and/or vascular fluid (Bynum and Porter 2006). It is proposed that 'mouth lobes' at the infracapitulum apex provide a physical seal around stylets when feeding, generating a vacuum and pumping action propagated by the pharynx and into the digestive tract. A distinct posture is adopted to enable this direct contact, the rear of the body inclined and supported by the infracapitulum and L1-L2 limbs. To guard against excess internal pressure as fluid enters the body, the small buccal cavity is connected to the oesophagus and a direct canal to the hindgut for excretion (Crooker et al. 2003).

Leaf chlorosis ensues at feeding sites, induced by lack of chlorophyll for photosynthesis and collapse of leaf mesophyll that prevents effective transpiration (Florida 2005). Chlorotic spots produce a yellow stippling on leaves, diagnostic of mite infection (Figure 2.2.4b,vA). Spider mites preferentially attack the lower surfaces of leaves closest to the ground, moving to the upper surface and higher up the plant as the population increases (Bynum and Porter 2006). As the colony grows and feeds more extensively, webbing is introduced to affix eggs and facilitate access to more host leaves (Figure 2.2.4b,vB), and spots coalesce to give yellow-grey blotches. *Tetranychus urticae* very quickly acquires resistance to all chemical acaricides, and although biological control species (e.g. *Phytoseiulus persimilis*, Figure 2.2.4b,vi) are available, defoliation and at least 20% loss of commercial crop yield is generally the outcome of any established infestation (Bynum and Porter 2006; Florida 2005).

2.2.5 Reproduction

i) Fertilisation

Prior to female ecdysis from her teliochrysalis, the male often adopts guarding behaviour (Figure 2.2.5,i), detecting a mature teliochrysalis by means of semiochemicals and/or dorsal lobe morphology within the chrysalis (Boudreaux 1963; Regev 1975; Regev 1980). Internal fertilisation generally occurs immediately after the teliochrysalis exuvia is cast off, and if multiple matings take place, the first mating appears to be the only effective one, fertilising eggs early in oogenesis and preventing further sperm entry (Boudreaux 1963). The fact that some females lay eggs that are all or partly unfertilised supports the proposition that successful fertilisation must take place shortly after ecdysis,

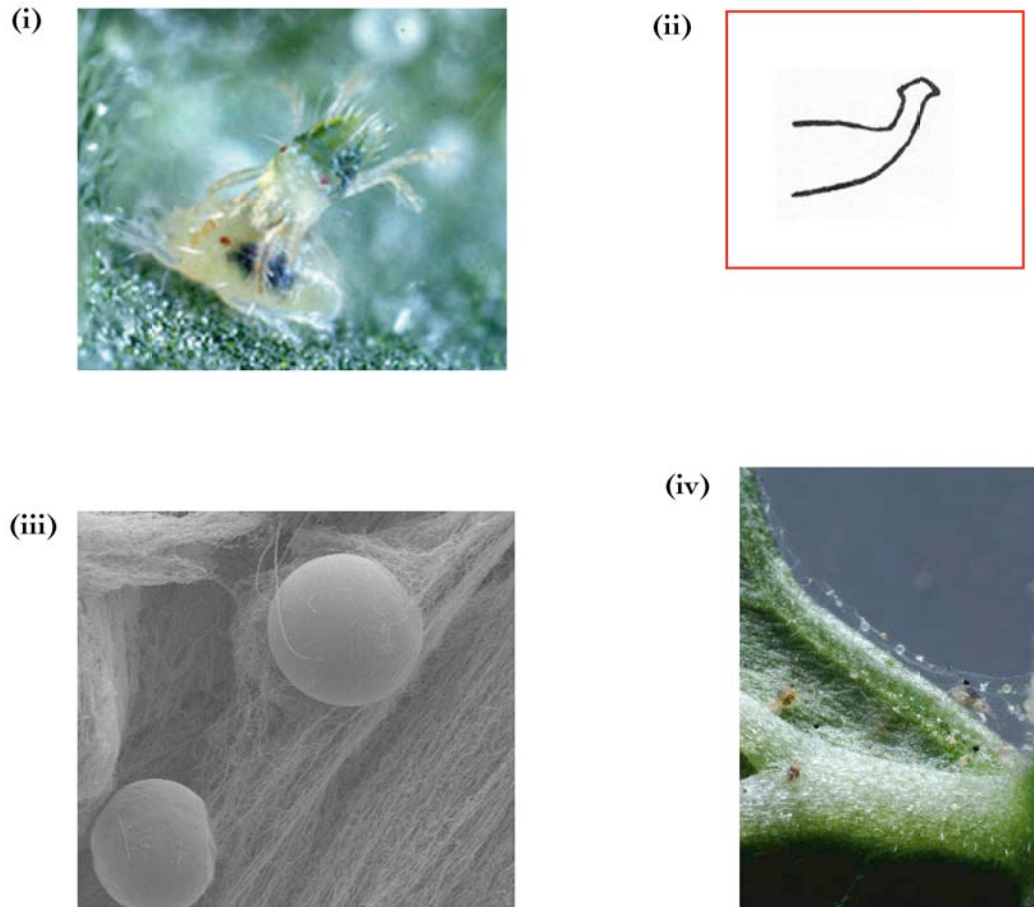


Figure 2.2.5 (i) Male guarding female deutochrysalis. (ii) Species-specific aedagus form in *Tetranychus urticae*. (iii) SEM spider mite eggs attached to webbing. (iv) Infested leaf, with reproductive population and mite eggs laid on leaf surface or web structure. Image (ii) from Flechtmann & Knihinicki (2002); SEM (iii) from Bielza (2005).

when oogenesis is in preliminary stages, and the oocytes presumably more penetrable. The male aedeagus is sclerotised to allow direct sperm transfer: within Chelicerata, this specialised internal mode of fertilisation is only found in tetranychid mites and Opiliones (Anderson 1973; Boudreaux 1963). The aedeagus is curved dorsally in all *Tetranychus* species, and its particular shape (Figure 2.2.5,ii) is diagnostic of *Tetranychus urticae* (Flechtmann and Knihinicki 2002). Unfertilised females lay eggs that develop into haploid males, and fertilised eggs develop into females. Such a haplo-diploid life cycle makes the species potentially amenable to genetic hybrid screens or transgenesis, as discussed previously in Chapter I regarding functional analysis of candidate genes.

ii) Egg laying and fecundity

Spherical, transparent eggs of 75µm average diameter are laid individually on the lower leaf surface and attached with fine silk threads (Figure 2.2.5,iii). When mite population density is high, eggs are laid above the leaf surface in a network of elevated webbing (Figure 2.2.5,iv). Eggs are impermeable, having a vitelline membrane, chorion and outermost waxy cuticle. Individual females may lay 60-100 eggs in a lifetime, commencing 1-2 days after emergence and laying fewer eggs if they are unfertilised. Eggs become more opaque and turn pale yellow-green during embryogenesis, with red pigmented larval eye spots (mentioned in previous sections) apparent approx. 10hrs before hatching.

Conclusions to Chapter II

i) *Tetranychus urticae* development: conserved or divergent?

In this Chapter we have seen that embryonic and post-embryonic development of *Tetranychus urticae* is characterised by features conserved among arthropods (e.g. labrum formation), conserved among chelicerates (e.g. pedipalp lobe formation) and more specifically among the Acariformes (e.g. single chorion layer, gnathosoma secondarily specialised for predation, specialised body segmentation, L4 limb repression, hexapod larval stage). However, *T. urticae* ontogeny also displays certain notable idiosyncrasies (e.g. posterior to anterior early germband differentiation, reduced cheliceral segment delay, loss of tritonymph stage, internal fertilisation, integument striation, sensory chaetotaxy), which

must be considered alongside more typical features when analysing associated gene expression patterns.

Conserved developmental features

The spider mite egg is typical of small, derived acarids in having a reduced structure, with outer waxy cuticle, single chorion layer and inner vitelline membrane; more basal acarids possess two chorion layers. A unitary yolk mass internal to the blastomere cell layer is found throughout the Acarida, as are scattered internal cells whose fate is uncertain but may derive from remnant cleavage, blastomere or endodermal cell populations. Both a posterior gastral groove in the germdisc and clear pedipalp lobes from earliest stages of limb bud development are typical of (synapomorphic for) the Arachnida in general. The fourth ambulatory limb pair (L4) does not develop fully during embryogenesis in any acarids, so the rudimentary limb buds and hexapod larval stage of *Tetranychus* also represent conserved features. The presence of a bilobed labrum supports an appendicular affinity for this conserved arthropod structure; as in the sarcoptiform *Archegozetes longesitosus*, the spider mite labrum appears long after limb development commences, is pronounced during limb elongation but is lost during germband contraction, as gnathal and stomodeal morphogenesis progress. I observed bilobed distal limb segments in all spider mite limbs at later stages, bilobed form reminiscent of other arthropods' 'biramous' limbs, or perhaps instead related to the complex sensory organogenesis characteristic of acarid tarsi. Once embryogenesis is complete, a typical acariform body plan is visible, characterised by anterior 'gnathosoma' (chelicerae and pedipalps specialised for sensory and feeding functions) and posterior 'idiosoma' (walking limbs covered by a dorsal carapace, fused to a very reduced opisthosoma). Body segmentation in *Tetranychus* includes a 'disjugal furrow' at the prosoma-opisthosomal boundary, ancestral to basal acarids (Opilioacariformes), and vestiges of a 'sejugal furrow' at the L2-L3 boundary, which is ancestral to Acariformes, dividing the body into anterior 'propodosoma' and posterior 'hysterosoma'. An extra leg segment - the genu - is present between femur and tibia in *Tetranychus* and many other acarids, and birefringent actinochitinous sensory setae on the limbs and dorsal carapace are a defining feature of the Acariformes (also known as 'Actinotrichida'), distinguishing them from the Parasitiformes.

Divergent developmental features

Mechanisms of germband formation and prosomal segmentation are remarkably diverse among acarids, and *Tetranychus urticae* is an unusual arthropod, exhibiting apparent posterior to anterior progression in early segment differentiation, reflected in the order of segmental groove appearance and direction of mesoderm differentiation. Delayed cheliceral segment specification is a synapomorphy for the Chelicerata, but in *Tetranychus urticae* this delay is negligible due to the speed and near-synchronicity of prosomal germband segmentation. Acarids typically display 5 opisthosomal

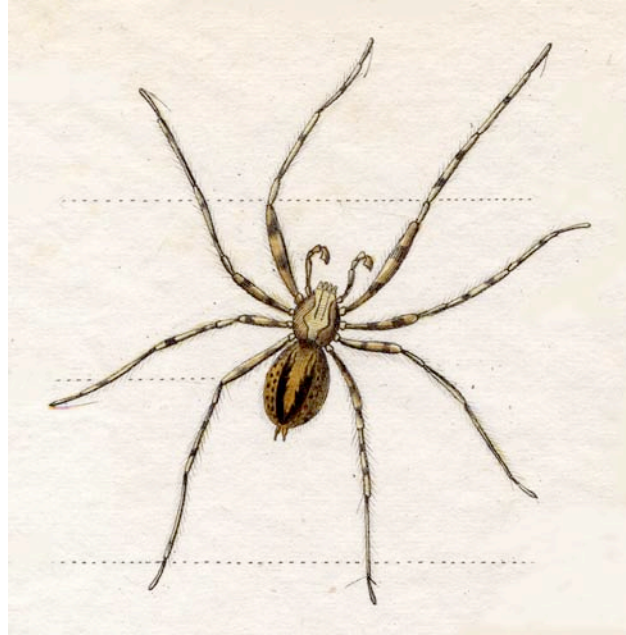
segments, but such segmentation is extremely reduced in tetranychids, such that opisthosomal segments are constricted or absent in *T. urticae* embryogenesis and no distinct opisthosoma is visible at all after hatching. Unique to the rapid life cycle of tetranychid spider mites is loss of the tritonymph stage, the deutonymph being immediately followed by a teliochrysalis from which a reproductive adult emerges. Although sexual dimorphism is common among acarids, the internal mode of fertilisation present in *Tetranychus urticae* is specific to the family Tetranychidae, and the shape of the male aedagus is diagnostic of, and strictly specific to the species *T. urticae*. Within the chelicerates as a whole, similar internal fertilisation only occurs in the Opiliones - presumably due to convergent evolution. Further defining features of the genus *Tetranychus* include the pattern of dorsal integument striation (diamond-shaped striae between two posterior setae, bilobed striae behind the stylophore), and the arrangement and form of sensory setae on the dorsal carapace and limbs. Finally, the name 'two-spotted' refers to bilateral accumulation of metabolites beneath the carapace, which although lightly pigmented green is largely transparent in *T. urticae* so that metabolic waste is visible externally.

ii) *Tetranychus* development and the breadth of comparative evo-devo

As mentioned previously, it is critical to interpretation of gene expression patterns to have a good understanding of development, as this allows accurate correlation of gene activity with particular structures, body parts or developmental events: the description of spider mite development in this chapter, distinguishing conserved, specialised and idiosyncratic features of ontogeny, satisfies that need. Furthermore, members of the major acariform sub-Order Trombidiformes, to which *Tetranychus urticae* belongs, have been little studied with regard to embryonic development (a recent detailed account of embryo development exists for the oribatid mite *Archegozetes longesitosus*, but it belongs to the other major acariform sub-Order, Sarcoptiformes). The validity and weight of evo-devo claims rest upon an appreciation of developmental genetics in a diverse range of taxa, and this Chapter's account of spider mite development provides the foundation to increase the breadth of our awareness to that end, providing a further chelicerate taxon for comparative study.

CHAPTER III

APPENDAGE SPECIFICATION IN *TETRANYCHUS URTICAE*: *DISTAL-LESS* AND *SP* GENES



Araignée domestique

Introduction to Chapter III

Tetranychus Distal-less

In investigating the molecular basis of limb specification in the spider mite, the gene *Distal-less* is of first importance, having been shown an early marker for appendage outgrowth throughout the animal kingdom, including in arthropod legs, polychaete parapodia, echinoderm tube feet and vertebrate limbs (Panganiban et al. 1997; Panganiban et al. 1994; Panganiban and Rubenstein 2002; Stopper and Wagner 2005). Within the arthropods, whose jointed appendages are directly homologous, *Dll* expression and function appears conserved in early limb specification along the antero-posterior axis, and in proximo-distal patterning during later outgrowth (Angelini and Kaufman 2005b; Cohen 1990; Cohen and Yurgens 1989; Mittmann and Scholtz 2001; Panganiban 2000; Panganiban et al. 1994; Panganiban and Rubenstein 2002; Prpic and Damen 2004; Prpic and Tautz 2003; Prpic et al. 2001; Schoppmeier and Damen 2001). In sections 3.1 and 3.2 of this chapter I report cloning of ~1.4Kb of a *Tetranychus Dll* (*Tu-Dll*) ortholog, confirmation of its orthology by Bayesian inference, and detection of *Tu-Dll* mRNA transcripts during embryogenesis by whole mount *in situ* hybridisation. The pattern of *Tu-Dll* gene expression is dynamic and indicates active roles in limb development, as well as in anterior neurogenesis; inferences regarding conservation of *Dll* gene function between *Tetranychus*, other arthropods and more disparate taxa, are discussed.

Tetranychus Sp genes

In addition to the well-known *Distal-less* gene, it has recently become clear that *Sp*¹³ genes play a critical role in limb development in both arthropods and vertebrates, suggesting a compelling level of genetic conservation and potential relevance to this study of appendage development in a chelicerate (Bouwman, 2002; Beermann, 2004; Estella, 2003; Triechel, 2003; Kawakami, 2004; Schöck, 1999). *Sp* family proteins are known to possess diverse roles in transcriptional regulation, with interactions mediated *via* diagnostic Sp, Buttonhead (Btd) and triple zinc finger (ZnF) domains (Athanihar et al. 1997; Bouwman and Philipsen 2002; Kadonaga et al. 1987). The *Sp* gene family is divided into two major sub-families: *Sp1/3/4*-type and *Sp8/9*-type genes, both sub-families hypothetically originating from an ancestral Urbilaterian gene pair formed by tandem duplication of a single ‘*Sp*-like’ gene (Figure 3.4.2c) (Kawakami et al. 2004). Distinct orthologs of *Sp8/9*-type and *Sp1/3/4*-type genes have been found in members of most major arthropod sub-phyyla, including myriapods (e.g. millipede

¹³ Nomenclature: *Sp* genes are known as ‘specificity protein’ genes, but the name originally derives from *sephacryl* and *phosphocellulose* columns used for their initial purification (Beermann 2004; Bouwman 2002; Kadonaga 1987).

Glomeris), crustaceans (e.g. *Parybale*) and insects (e.g. beetle *Tribolium*, flies *Clogmia* and *Drosophila*); thus far, chelicerates have not been surveyed (Wimmer, 2005; Janssen, 2005; Beermann, 2004; Wimmer, 1996; Schöck, 1999). The *Tribolium castaneum* *Sp8/9* ortholog, *Tc-Sp8*, seems to be involved in controlling limb allometry, and work in *Drosophila* indicates that a recently duplicated *Sp8/bt¹⁴* gene pair redundantly and positively regulates *Dll*; when ectopically expressed in dorsal tissue, *Sp8/bt¹⁴* can independently activate a full genetic network (*wg*, *dp¹⁴* etc.) sufficient for ventral limb specification (Beermann et al. 2004; Estella et al. 2003). In section 3.4 of this chapter I present preliminary data regarding *Sp* orthologs in *Tetranychus*, which were sought with a view to assessing the involvement of *Sp* genes in limb specification and development outside the insects, broadening the analysis to include a basal, chelicerate arthropod species. By degenerate PCR screening, I cloned short (~180bp) fragments of two spider mite *Sp* genes, shown by sequence analysis and Bayesian inference to be putative *Sp8/9* and *Sp1/3/4* orthologs. Attempted *in situ* hybridisation with small ssRNA probes was unsuccessful, most probably due to sub-optimal probe size; limited discussion is given regarding predicted *Sp* expression.

¹⁴ In referring to the *Drosophila melanogaster* *Sp8/bt¹⁴* gene pair, I have re-named *Sp8* to put it in line with current understanding of *Sp* gene orthologies. Original literature refers to '*D-Sp1*' and '*bt¹⁴*' genes, whereas '*D-Sp1*' is in fact an *Sp8* ortholog (Wimmer et al. 1996). See also Figure 3.4.2c.

3.1 Cloning of the *Tetranychus urticae Distal-less* gene

By degenerate PCR and subsequent inverse PCR I recovered almost 1.4Kb (1,358bp) of a *Tetranychus urticae Dll* ortholog (*Tu-Dll*), including 1140bp CDS and 216bp 3' UTR. A putative *Tu-Dll* EST sequence (885bp) became available to me from Agriculture Canada during my PhD, identical to the *Tu-Dll* gene reported here. Although the 1.4Kb *Tu-Dll* sequence lacks complete 5' CDS, the fragment proved adequate for the purpose of reporting mRNA transcription patterns *in situ* on whole mount embryos (for which, see section 3.3).

3.1.1 Degenerate PCR screening

A ~140bp part of the *Tetranychus Dll* homeobox was amplified by degenerate RT-PCR on total RNA and identified by BLAST sequence similarity search (Figure 3.1.1a,i). Amplification with primers PanDll F1 (5'-CGT AAR CCN MGN CAN ATH TA-3'), PanDllF2 (5'-CGT AAG CCG MGN CAN ATH TA-3') and PanDll R1 (5'-ACG RTT YTG RAA CCA DAT YTT-3') yielded bands of the expected size. Forward primers target conserved N' Dll/Dlx family homeodomain residues KPRTIY and reverse primers the more generic C' homeodomain residues KIWFQN (see Appendix 2.1 for further primer details). The degenerate PCR program (degDll; see Materials and Methods) was based on that used by Panganiban et al. (1995) to amplify *Dll* homologs from diverse metazoan taxa (Panganiban et al. 1995). Initial low temperature (43°C) cycles provide low stringency for primer binding within the broad *Dll/Dlx* gene family, and later 50°C cycles rapidly amplify *Dll/Dlx* target homeobox fragments obtained. As well as spider mite *Dll*, a few non-target homeobox genes were amplified, however, including putative *Msx*, *Aristal-less*/Zn finger gene, and a 144bp *Hox* gene fragment most resembling a posterior class *abdominal-B/Hox-D9* gene (Figure 3.1.1a,ii-iv). *Msx* genes encode transcription factors that can heterodimerise with Dlx *via* highly conserved homeodomains, and are thought to have roles in ectoderm-mesenchyme interactions in vertebrate limb buds, craniofacial processes and tooth buds as well as echinoderm oral ectoderm and mesoderm patterning (Lallemand et al. 2005; Panganiban and Rubenstein 2002; Wilson et al. 2005). *Aristal-less* is a homeobox gene and downstream target of Dll, expressed distally during proximo-distal limb patterning, while the canonical *Hox* genes (*abdB/D9* in this case) act upstream of *Dll* in A/P axial patterning, sharing with Dll conserved residues at the C' terminus of the homeodomain, but differing in sequence at the N' Hox-diagnostic 'LELEKEF' domain (Campbell and Tomlinson 1998; de Rosa et al. 1999).

(i) *Tu-Dll*

01/01
AAG CCG AGA ACG ATT TAT TCA AGT ATT CAA CTT CAA CAG CTT AAC AGA CGA TTT CAA AGG
lys pro arg thr ile tyr ser ser ile gln leu gln gln leu asn arg arg phe gln arg

ACA CAG TAT TTA GCT TTA CCC GAA AGA GCT GAG TTA GCA GCA TCA CTA GGA TTG ACT CAA
Thr gln tyr leu ala leu pro glu arg ala glu leu ala ala ser leu gly leu thr gln
46/138
ACA CAG GTT AAG ATA TGG
Thr Gln val lys ile trp

(ii) *Tu-Msx*

01/01
CGT AAG CCG AGA ACG ATC TAC AAC TCA ACA GTT ATT AGC GCT GGA AAG AAA GTT TCA GTC
arg lys pro arg thr ile tyr asn ser thr val ile ser ala gly lys lys val ser val

CAA ACA GTA CTC TTA TCA ATT GCC GAA AGA GCA GAG TTT TCA ACA TCT TTA AAC TTA ACG
gln thr val leu leu ser ile ala glu arg ala glu phe ser thr ser leu asn leu thr
51/153
GAA ACT CAG GTT AAG ATC TGG TTC CAA AAC CGT AAT
glu thr gln val lys ile trp phe gln asn arg asn

(iii) *Tu-Al/zinc finger*

01/01
ATT CGA TTT TGG ACC GTC GGT TCT GGA ACC ACT CCT GGT AAA CGT GCC AAT AGA ACT CGT
ile arg phe trp thr val gly ser gly thr thr pro gly lys arg ala asn arg thr arg

TTC ACC GAT TAT CAG ATT AAA GTA TTG CAA GAA TTT TTT GAA TCA AAT GCA TAC CCT AAA
phe thr asp tyr gln ile lys val leu gln glu phe phe glu ser asn ala tyr pro lys

GAT GAT GAT CTT GAA TAT CTT TCC AAG CTT CTT AAC CTT TCA CCT CGT GTT ATT GTT GTT
asp asp asp leu glu tyr leu ser lys leu leu asn leu ser pro arg val ile val val
70/210
TGG TTC CAA AAC CGA CGG TCC AAA
trp phe gln asn arg arg ser lys

(iv) *Tu-abdB/boxD9*

01/01
ATT CGT AAA CCG AGG ACG ATA TAT CAA ACG GTT GAA CTG GAA AAA GAG TTT ACT GTC AAT
ile arg lys pro arg thr ile tyr gln thr val glu leu glu lys glu phe thr val asn

TTC TAT GTA ACC AAA CAA CGT CGA TTT GAA CTG TCC CGA GCT TTA GGA CTT TCA GAG AGA
phe tyr val thr lys gln arg phe glu leu ser arg ala leu gly leu ser glu arg
48/144
CAG GTT AAG ATC TGG TTT CAA AAC CGT
Gln val lys ile trp phe gln asn arg

Figure 3.1.1a Nucleotide sequences and amino acid translations for PCR screening products obtained using degenerate *Dll* primers. Grey: conserved primer target sequences. Underline in (iv): conserved *Hox* 5' LELEKEF homeodomain motif VELEKEF. Numbers refer to nucleotide and amino acid sequence positions.

3.1.2 Inverse PCR and EST data

Relative to the original degenerate PCR fragment, inverse PCR provided a further 1112bp in the 3' direction (past the stop codon into 216bp 3' UTR), and a further 106bp 5' coding sequence. See Figure 3.1.2a for complete nucleotide sequence, amino acid translation and an indication of the extent of the gene also represented in the *Tu-Dll* EST. Successful amplification outwards from known *Tu-Dll* sequence was achieved using gene-specific primers L208, L212 and U680 on genomic DNA digested with BamHI or XhoI before circularisation (see Appendix 2.1 for primer sequences). Having converted circularised DNA-derived sequence data into genomic orientation (Figure 3.1.2b) and compared the new 1,358bp total sequence with the EST data, I calculated that of 1218bp obtained by inverse PCR, 474bp were novel *Tu-Dll* data.

3.1.3 Phylogenetic analysis: Bayesian inference

I carried out Bayesian analysis of a subset of *Tu-Dll* coding nucleotides, including the conserved homeodomain and a few residues outside it (see Figure 3.1.3a for multiple sequence alignment); this dataset comprises that fraction of the gene unambiguously alignable as homologous amino acid sites (Hughes and Kaufman 2002b; Mann and Carroll 2002; Merlo et al. 2000; Panganiban 2000). *Msx* orthologs and several other homeobox-containing genes (i.e. *eve*, *En*, *Ubx*, *abdB*) were added to an alignment of *Dll/Dlx* orthologs, for two reasons: Firstly, *Msx* genes are among those most closely related to *Dll/Dlx* and hence should root any hypothetical *Dll/Dlx* clade effectively, and secondly, other homeobox genes provide more distant outgroups and means to confirm the proposed *Dll/Dlx* + *Msx* relationship (Prpic 2004b).

i) Full dataset analysis

Bayesian consensus (Figure 3.1.3b tree) places *Tetranychus Distal-less* within a maximal-support *Dll/Dlx* clade comprising arthropod and chordate orthologs. *Hydra Dll-related* is basal to the *Dll/Dlx* clade (0.94 posterior probability), and similarly high probabilities support a sister group relationship between the *Hydra* + *Dll/Dlx* clade and *Msx* genes. Phylogenetic resolution within the *Dll/Dlx* clade itself is poor and deviates from the pattern represented in most species phylogenies with respect to both chordates and arthropods (Blaxter 2001; Cameron et al. 2000; Cook et al. 2001; Cracraft and Donoghue 2004; Delsuc et al. 2006; Nardi et al. 2003a; Nardi et al. 2003b; Niedert et al. 2001). Presumably there is inadequate phylogenetic signal in the analysed character set for full resolution, or branching patterns were affected by variability in relative rates of *Dll/Dlx* gene evolution between species. It is not unusual that a pattern of gene evolution differs from that put forward to represent

1/1 31/11 61/21

GCA CGA GGC ATC ACC TCA TTT CCA GCC TGT CCA ACA CCT CCA AGA GAC GAG AAA CCA ACA TTG GAG GAA ATA
ala arg gly ile thr ser phe pro ala cys pro thr pro pro arg asp glu lys pro thr leu glu glu ile
91/31 121/41

TCT CGT GTA AAT GGT AAA AAT AAA AAA AAC TGG AAA CCT CGT ACA ATT TAT TCA AGT ATT CAA CTT CAA CAG
ser arg val asn gly lys asn lys lys asn trp lys pro arg thr ile tyr ser ser ile gln leu gln gln
151/51 181/61 211/71

CTT AAC AGA CGA TTT CAA AGG ACA CAG TAT TTA GCT TTA CCC GAA AGA GCT GAG TTA GCA GCA TCA CTA GGA
leu asn arg arg phe gln arg thr gln tyr leu ala leu pro glu arg ala glu leu ala ala ser leu gly
241/81 271/91

TTG ACT CAA ACA CAG GTT AAA ATA TGG CTA CAA AAT AAA CGT TCA AAA AAC AAG AAA ATG CAA AAA GCT CAG
leu thr gln thr gln val lys ile trp leu gln asn lys arg ser lys asn lys lys met gln lys ala gln
301/101 331/111

GAA GCT GTG AAT GGA GGT GGA CAA GTG AAT GGT TCA CAA GGT GTT GCC TGT GGT ACT GGT GGT GGT CGA CGA
glu ala val asn gly gly gly gln val asn gly ser gln gly val ala cys gly thr gly gly gly arg arg
361/121 391/131 421/141

GGT CGA GGA GGT AAT CAA GGT CAA GGG CAG AAT CAA TCG CAG CAA CAG CAA CAA GCA CAA CAA CAG GCC CAA
gly arg gly gly asn gln gln gly gln asn gln ser gln gln gln ala gln gln gln ala gln
451/151 481/161

CAT CAG CAG CAG CAA CAG CAG CAG CAA CAA CAC CAA CAA CAA CAC CAA CAG CAC CAA CAA CAA CAG CAA
his gln gln gln gln gln gln gln gln gln his gln gln gln his gln gln his gln gln gln gln
511/171 541/181 571/191

CAA GCT CAA GTC CAA CAA ACT CTT TCA AAT CAA CAG CAA TCA CTT GAT ACT CAA CCA GGA GTT GGT CTA TCC
gln ala gln val gln gln thr leu ser asn gln gln gln ser leu asp thr gln pro gly val gly leu ser
601/201 631/211

TCG GGT TCA CCT TTC ATC AAA GGT GAA GGT TAC ATT CCT CAA CAT TCC CCT GAG GTA CCA AGT GAA AGC CAT
ser gly ser pro phe ile lys gly glu gly tyr ile pro gln his ser pro glu val pro ser glu ser his
661/221 691/231

ACA CCA TTA CAT TCA TCA CTT GGT CCA AAT CCG GGT TCC AAT GGT GTC AAC TCA AAT GGC CCT GGT GCA CCT
thr pro leu his ser ser leu gly pro asn pro gly ser asn gly val asn ser asn gly pro gly ala pro
721/241 751/251 781/261

GTT AAT AGT ATC GGT AAT GCA AAC CTA GGA AAC TCC CTT GTA TCC TCG ATA GTC TCC TCT TCA TCT TCT TCT
val asn ser ile gly asn ala asn leu gly asn ser leu val ser ser ile val ser ser ser ser ser
811/271 841/281

TCC TCT TCC TCA ACA GCC TCT GGT CCT GTT CTT ACT CGT TCC CCG GTT CTT CCG GTT TAC TCT GAA TCA CCA
ser ser ser ser thr ala ser gly pro val leu thr arg ser pro val leu pro val tyr ser glu ser pro
871/291 901/301 931/311

CTT AAC TCA TCA CAA ACG ATT GCT CGA AGT TCA CCT GGA ACC ATT CAT TCA CCT TCC CAC ACT CCC AAC CAA
leu asn ser ser gln thr ile ala arg ser ser pro gly thr ile his ser pro ser his thr pro asn gln
961/321 991/331

TGG TCC TCA ATG GGT CAG ATA ACA CCA AAA CAA GAG ATC GGT CTT GGC CTG GGT TCC TCA TCG TCC TCC TCA
trp ser ser met gly gln ile thr pro lys gln glu ile gly leu gly leu gly ser ser ser ser ser
1021/341 1051/351

TCG TCA ACC TCT TCC TCT TCT GTG GGT GGA GGC AGT GCA CTA GGT AAC ACT GGA CAT CCT CCT TAT CCA CCT
ser ser thr ser ser ser val gly gly gly ser ala leu gly asn thr gly his pro pro tyr pro pro
1081/361 1111/371 1141

CCA CAT CCA TCA ATG TAC AGC AAT AGA AAA GTC GAA TTT TTC ACC AAA AAA AAA GAT TAA ATT TTT ATC GCA
pro his pro ser met tyr ser asn arg lys val glu phe phe thr lys lys asp ⓧ
1171 1201

ATT GTG TTA AAG TAT GAA AGC ATT AAA GTA GTT GAA GTA TTA AGA AAG TGA AAA GTT TCA AGA TTT TAA AAA
1231 1261 1291

ACA ATG TTT TCG ATC AAA AGA GTC ATG AAA AAA TGA GTT GAA AAA TGG GCA AAG TTT CAT TTT CAG TAC TTT
1321 1351

AGC ACT TGA AAC GTT TTC TAA TCA AAA TTT CAT TCT TTT CTG ATT TTC TTT CAT GAT TGT

Figure 3.1.2a *Tetranychus-Dll* sequence data obtained by inverse PCR; sequence also available from EST in grey. **Yellow**: homeodomain, with TIY (thr-ile-tyr) and KIW (lys-ile-trp) 5' and 3' Dlx motifs in blue (Panganiban & Rubenstein, 2002; de Rosa et al., 1999). Stop codon (TAA) is indicated by ⓧ.

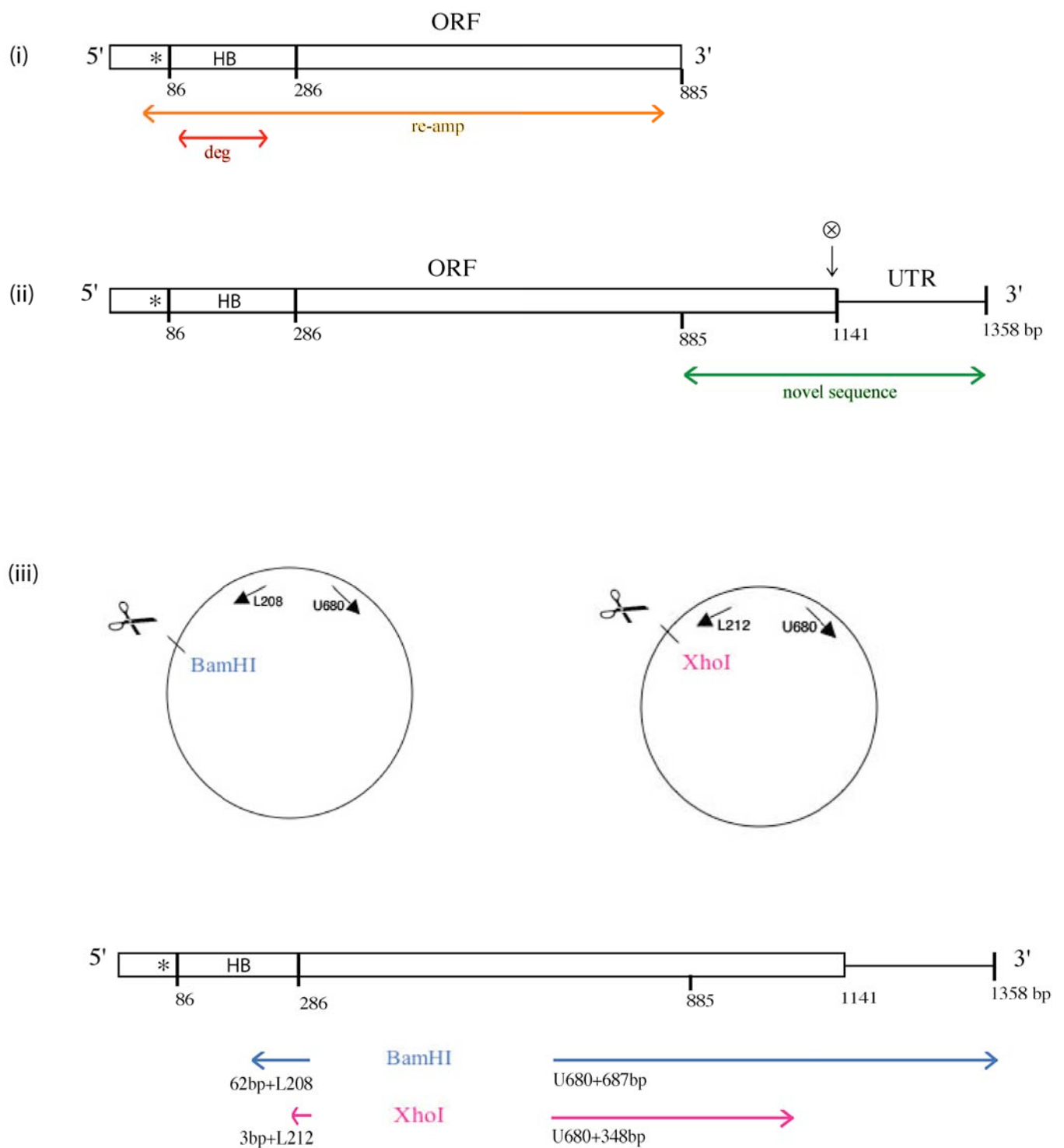


Figure 3.1.2b (i) Schematic representation, to scale, of Tu-DII gene structure determined from EST sequence data. * = LEE motif (chelicerate-specific), HD = homeobox, ORF = open reading frame (coding sequence), UTR = untranslated region. The orange arrow delimits the region re-amplified from EST-bearing plasmid DNA using U41 + L865 primers. The red arrow represents the extent of the original Tu-DII fragment cloned by degenerate PCR. **(ii)** Scale representation of 1358bp Tu-DII fragment obtained by inverse PCR. The green arrow shows novel sequence relative to the EST, of 433bp. ⊗ = TTA stop codon. **(iii)** Schematic illustrations of iPCR results in circular-digest, and interpreted-genomic orientations. DNA circles are shown (not to scale) with outward sequencing primers and an approximation of the relative site of restriction by both BamHI and XhoI. Coloured arrows indicate sequencing from primers in the known region up to the BamHI/XhoI cut site. Sequencing from the BamHI (blue) circularised digest produced 740bp outlie the known region (62bp downstream, 687bp upstream), and from the XhoI (pink) 351bp were sequenced (3bp upstream, 348bp downstream). The BamHI DNA circle was 53% larger than the XhoI, and yielded novel sequence to append to EST data, although lack of an AATAAA ribosome-cleavage initiation site or terminal poly-adenylated 'tail' suggests that 3' mRNA data is not entirely complete.

	N'							C'
Tetranychus Dll	WKPRTIYS	SIQLQQLNRR	FORTQYLALP	ERAELAASLG	LTQTQVKIWL	QNKRSKNKKM	QKA	
Cupiennius Dll	R.....	.L.....F	..R...Y...	L..	Arthropoda
Glomeris Dll	R.....	.L.....F	..R...Y...	M..	
Oncopeltus Dll	R.....	.L.....F	..R...Y...	M..	
Schistocerca DllL.....F	..R...Y...	M..	
Tribolium Dll	R.....	.L.....F	..R...Y...	M..	
Bicyclus Dll	R.....	.L.....F	..R...Y...	M..	
Junonia Dll	R.....	.L.....F	..R...Y...	M..	
Anopheles Dll	R.....	.L.....F	..R...Y...	M..	
Manduca Dll	R.....	.L.....F	..R...Y...	M..	
Drosophila DllB	R.....	.L.....F	..R...Y...	M..	
Drosophila DllA	R.....	.L.....F	..R...Y...	M..	Lower + Hemi- Chordata
Saccoglossus Dlx	R.....	.L.....F	..R...C...	L.Q	
Oikopleura Dlx	R.....	.L...E..K.	.NQ.....F	..R...Y..I	M.Q	
Ptychodera Dlx	R.....	.L.....F	..R...Y..V	L.Q	
Ciona DllC	R.....	.L...A...	.Q.....T..F	..R...C..L	M.Q	
Ciona DllA	R.....T	.Y...V...	V...I...F	..R...NY..L	L.Q	
Branchiostoma Dll	R.....T	.F.....Q..F	..R...Y..L	M.Q	
Homo Dlx1	R.....	.L...A...	.Q.....F	..R...F..L	M.Q	
Homo Dlx5	R.....	.F...AA.Q..	.K.....F	..R...I..I	M.N	
Homo Dlx6	R.....	.L...A..H.	.Q.....F	..R...F..L	L.Q	
Homo Dlx2	R.....	.F...AA.Q..	.K.....F	..R...F...W	S	Craniata
Homo Dlx4	R.....	.L...H..Q.	.H.....	...Q...Q..F	..R...Y..L	L.Q	
Homo Dlx4a	R.....	.L...H..Q.	.H.....	...Q...Q..F	..R...Y..L	L.Q	
Homo Dlx3	R.....	.Y...AA.Q..	.KA.....	...Q...Q..F	..R...F..L	Y.N	
Homo Dlx7	R.....	.L...H..Q.	.H.....	...Q...Q..F	..R...Y..L	L.Q	
Mus Dlx1	R.....	.L...A...	.Q.....F	..R...F..L	M.Q	
Mus Dlx2	R.....	.F...AA.Q..	.K.....F	..R...F...W	S	
Mus Dlx6	R.....	.L...A..H.	.Q.....F	..R...F..L	L.Q	
Mus Dlx5	R.....	.F...AA.Q..	.K.....F	..R...I..I	M.N	
Mus Dlx5alpha	R.....	.F...AA.Q..	.K.....F	..R...I..I	M.N	
Mus Dlx3	R.....	.Y...AA.Q..	.KA.....	...Q...Q..F	..R...F..L	Y.N	Cnidaria
Mus Dlx7	R.....	.L...H..Q.	.H.....	...Q...Q..F	..R...Y..L	L.Q	
Mus Dlx4	R.....	.L...H.DQ.	.H.....	...Q...Q..F	..R...Y..L	L.Q	
Danio Dlx1a	R.....	.L...A...	.Q.....F	..R...F..L	M.Q	
Danio Dlx5a	R.....	.F...AA.Q..	.N.....F	..R...L..I	M.N	
Danio Dlx6a	R.....	.L...A..H.	.Q.....F	..R...F..L	L.Q	
Danio Dlx2a	R.....	.F...AA.Q..	.K.....F	..R...F..L	W.S	
Danio Dlx4b	R.....	.V...A.HQ.	.Q.....	...D...K..F	..R...Y..I	M.H	
Petromyzon DlxD	R.....	.L...A...	.Q.....F	..R...Y..L	M..	
Xenopus Dlx4	R.....	.F...AA.Q..	.K.....	V.....F	..R...F...W	S	
Hydra Dll-r	R.....FT	.H...RE...S	.E...H..S..HG..I...F	..R...F..I	I..	
Mus Msx2	R....PFT	TS...LA.E.K	.RQK...SIA	...FSS..N	..E.....F	..R.A.A.RL	.E.	Msx proteins
Mus Msx1	R....PFT	TA...LA.E.K	.RQK...SIA	...FSS..S	..E.....F	..R.A.A.RL	.E.	
Mus Msx3	R....PFT	TA...LA.E.K	.HQK...SIA	...FSS..S	..E.....F	..R.A.A.RL	.E.	
Homo Msx4	R....PFT	TS...LA.E.K	.RQK...SIA	...FSS..N	..E.....F	..R.A.A.RL	.E.	
Homo Msx1	R....PFT	TA...LA.E.K	.RQK...SIA	...FSS..S	..E.....F	..R.A.A.RL	.E.	
Oryzias Msx4	R....PFT	TS...LA.E.K	.RQK...SM.	...FSS..T	..E.....F	---	
Cupiennius eve	RRY...AFT	RE...AR.EKE	.M.EN.VSR	R.C...TA.N	.PESTI.V.F	..R.M.D.RQ	RMS	other HD proteins
Cupiennius En1	KR...AFT	AD...SR.KHE	..ENR..TER	R.QD...KD.Q	.NES.I...F	..R.A.L..A	SGQ	
Cupiennius En2	RRC...AFT	AH.VSR.RNE	...SE...SEA	G.R...QE.Q	..EA.....F	..A.A.RR.N	CGR	
Cupiennius Ubx1	RRG.QT.T	RY.TLE.EKE	.HTNH..TRR	R.I.M.HA.C	..ER.I...F	..R.M.L..E	IQ.	
Cupiennius Ubx2	RRG.QT.T	RY.TLE.EKE	.HTNH..TRR	R.I.M.H..C	..ER.I...F	..R.M.L..E	AQ.	
Cupiennius abdB	R.K.KP..	KF.TLE.EKE	.LFNA.VSKQ	K.W...RN.N	..ER...F	..R.M.S..T	SQR	

Figure 3.1.3a Multiple alignment of Dll/Dlx amino acid sequences for taxa used in phylogenetic analysis; Msx and other homeodomain proteins (eve, Hox) are to serve as outgroups. The alignment of 61 amino acids was calculated in PHYLIP v.3.6 using Jones-Taylor-Thornton model parameters. Dots (.): identical sites relative to *Tetranychus* Dll. Gaps (-): missing data. Grey: strict homeodomain (59aa, 177bp). Yellow: N' and C' terminal motifs characteristic of Dlx and HD proteins respectively.

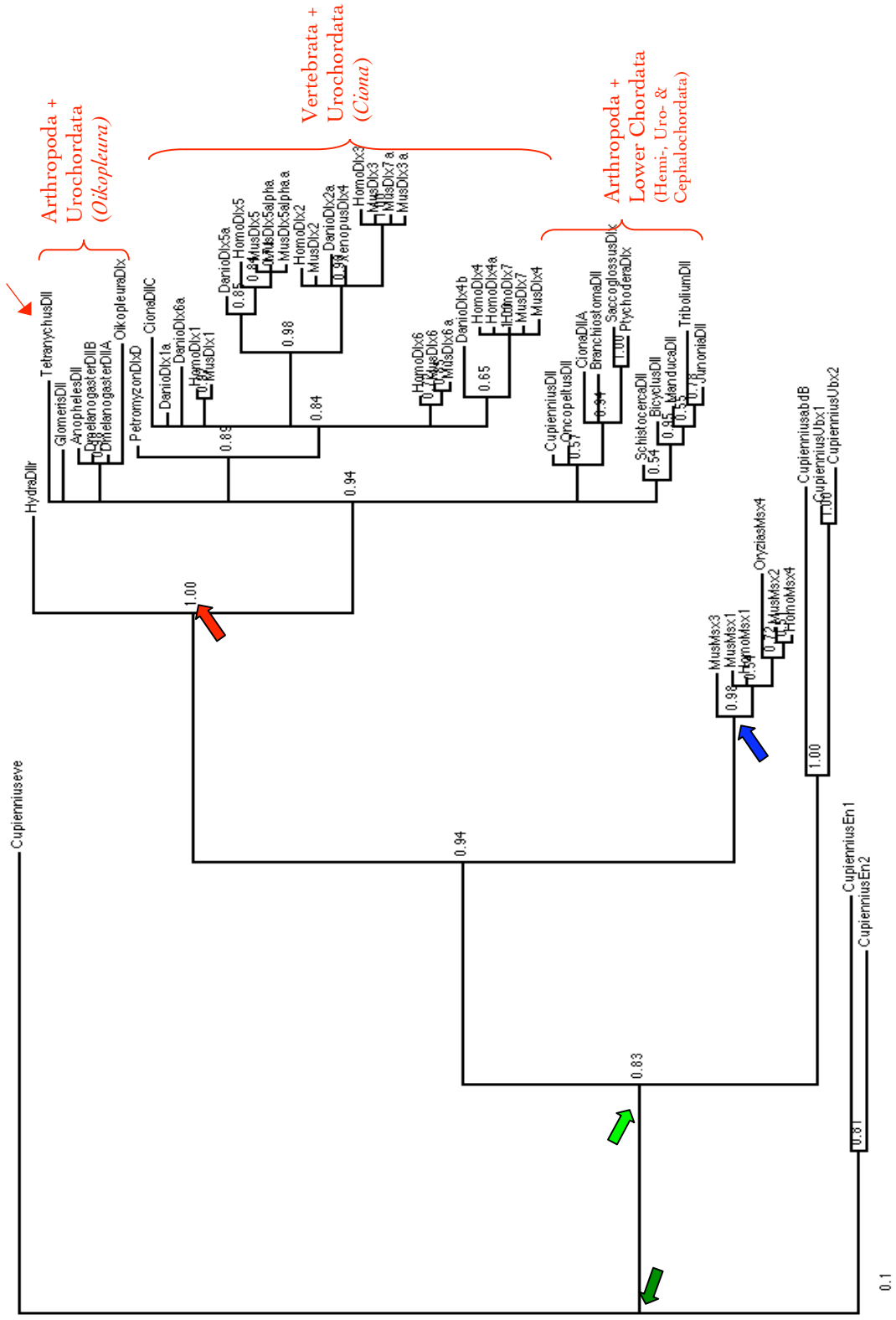


Figure 3.1.3b Bayesian consensus tree for *Tu-Dll* and related coding nucleotide sequences. *Tetranychus Dital-lew* (red arrow) is part of a *Dll/Dlx* clade (red filled arrow) with full statistical support and distinct from a near-maximally supported (0.94 posterior probability) *Mav* clade (blue filled arrow) and an 83% supported outgroup comprised of spider *Hav*-*En* genes (green filled arrow), spider *eve* falling just outside the primary outgroup clade. Within the *Dll/Dlx* clade, *Tetranychus Dll* falls within an 'Arthropoda + Hemichordata' group. Outgroup *Hav*, *En* and *eve* sequences are relatively long branches, which may explain incomplete resolution of *En/eve* branching (dark green filled arrow) relative to the *Hav* genes.

species evolution, however, effects such as long branch attraction may also distort the apparent relationships between gene sequences (c.f. Chapter VIII section 8.3.1). *Oikopleura*-*Dlx*, and to a lesser extent, *Tetranynchus*-*Dll*, form relatively long branches, potentially explaining why the urochordate *Oikopleura* groups with higher insects (Diptera) in Figure 3.1.3b, rather than with other urochordates (e.g. *Ciona*) or chordates in general.

ii) Reduced dataset analysis

In order to characterise relationships within the *Tu-Dll* clade without any negative effects from the necessarily longer branches of more distant outgroup sequences, I subjected a *Dll/Dlx*-only dataset to Bayesian analysis (Figure 3.1.3c). In this case, posterior probability support values were generally lower than those obtained with the larger dataset, but *Tetranynchus Dll* bears closer affinity to spider and myriapod sequences than to the more derived insects, and the arthropods form a monophyletic group, more in keeping with accepted ideas on arthropod relationships (Blaxter 2001; Cook et al. 2001; Moore 2001; Nardi et al. 2003a; Nardi et al. 2003b). Although deuterostome relationships are also clearer, the topology diverges from standard assumptions that hemichordates are most ancestral, with urochordates, cephalochordates, craniates and finally vertebrates diverging from the evolving lineage Chordata (Cameron et al. 2000; Cracraft and Donoghue 2004; Niedert et al. 2001). However, an impressive recent attempt to resolve deuterostome relationships by molecular phylogenetic methods solidly supports urochordates (e.g. *Ciona*, *Oikopleura*), and not cephalochordates, as most closely related to craniates (Delsuc et al. 2006): the *Dll/Dlx* gene topology concurs with this second scenario (Figure 3.1.3c). Other somewhat atypical aspects of the topology may reflect lineage-specific variation in relative rates of molecular evolution, or perhaps the long branch *Oikopleura-Dlx* disrupting an optimal phylogeny with respect to deuterostome members.

3.1.4 Digoxigenin-labelled mRNA probe synthesis and efficacy

Single stranded RNA (ssRNA) probes labelled with digoxigenin-alkaline phosphatase (DIG-AP) were synthesised after plasmid purification of *TuDll* inserts in Promega pGEM®-T Easy vector, inserts derived from either EST or PCR amplified material (Figure 3.1.4a, Figure 3.1.4b). The complete *Tu-Dll* insert was determined to be ~1.4Kb long, and sections of open reading frame ranging from 689bp to 804bp in length were amplified from genomic DNA, cDNA or EST plasmid template, by means of gene-specific primers (Figure 3.1.4a), insert sizes being confirmed after cloning by restriction enzyme (RE) digestion. Depending on specific insert orientation in pGEM®-T Easy vector and whether I was synthesising sense(+) or anti-sense(-) ssRNA, distinct digestion/polymerase

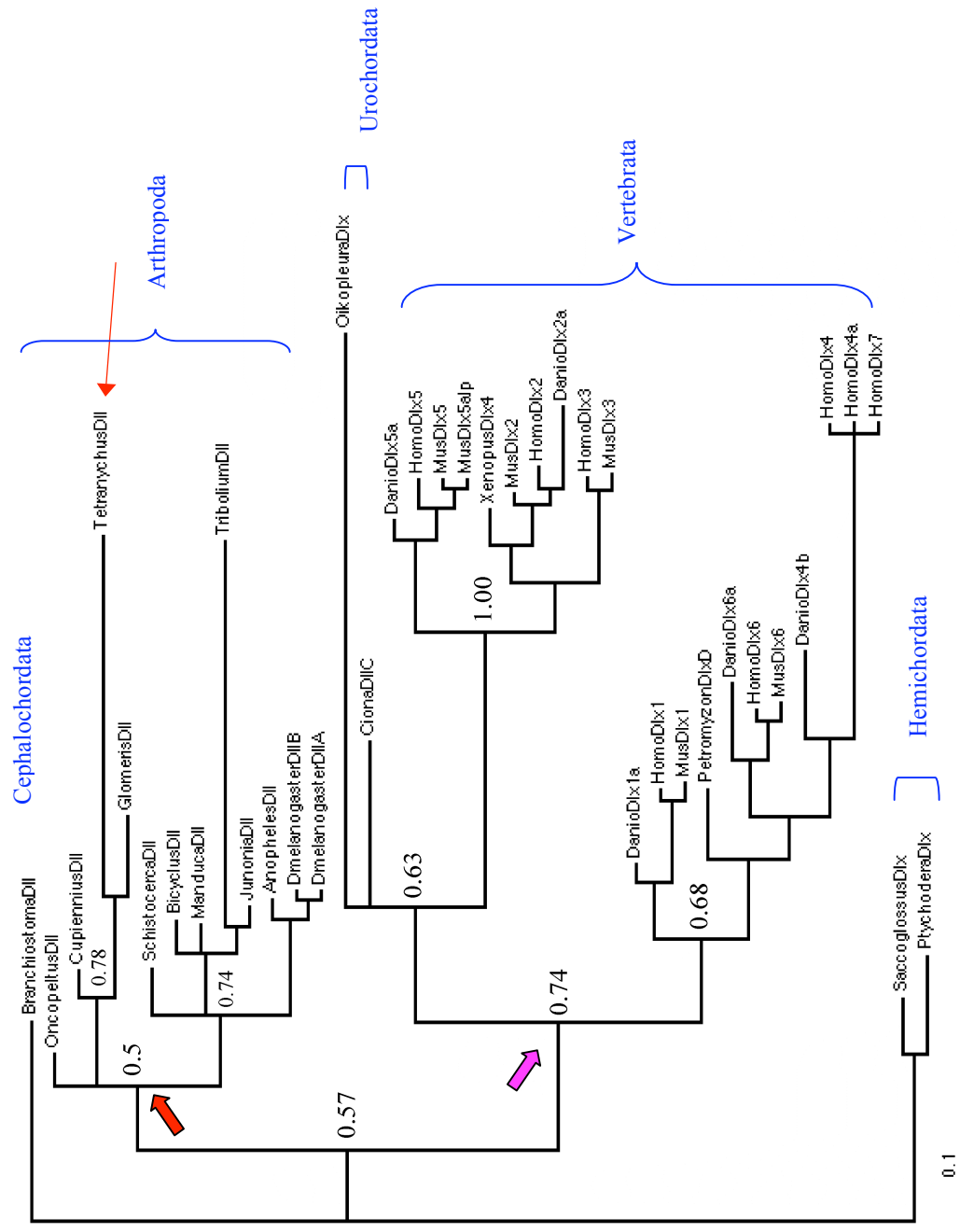


Figure 3.1.3c Bayesian consensus tree for *Tl-Dll* within a *Dll/Dlx* dataset. Arthropods (red filled arrow) form a separate clade from the chordates (pink filled arrow), and within the arthropods myriapods and chelicerates - including *Tl-Dll* (red arrow) - are separate from insects. With respect to *Dll-Dlx* sequence evolution, the cephalochordates and hemichordates seem to retain a more ancestral state than the vertebrates and urochordates; this reflects chordate relationships recently published in Delsuc et al. (2006). Long branches within the tree include *Oikopleura* and to a lesser extent *Tribolium* and *Tetraodon*.

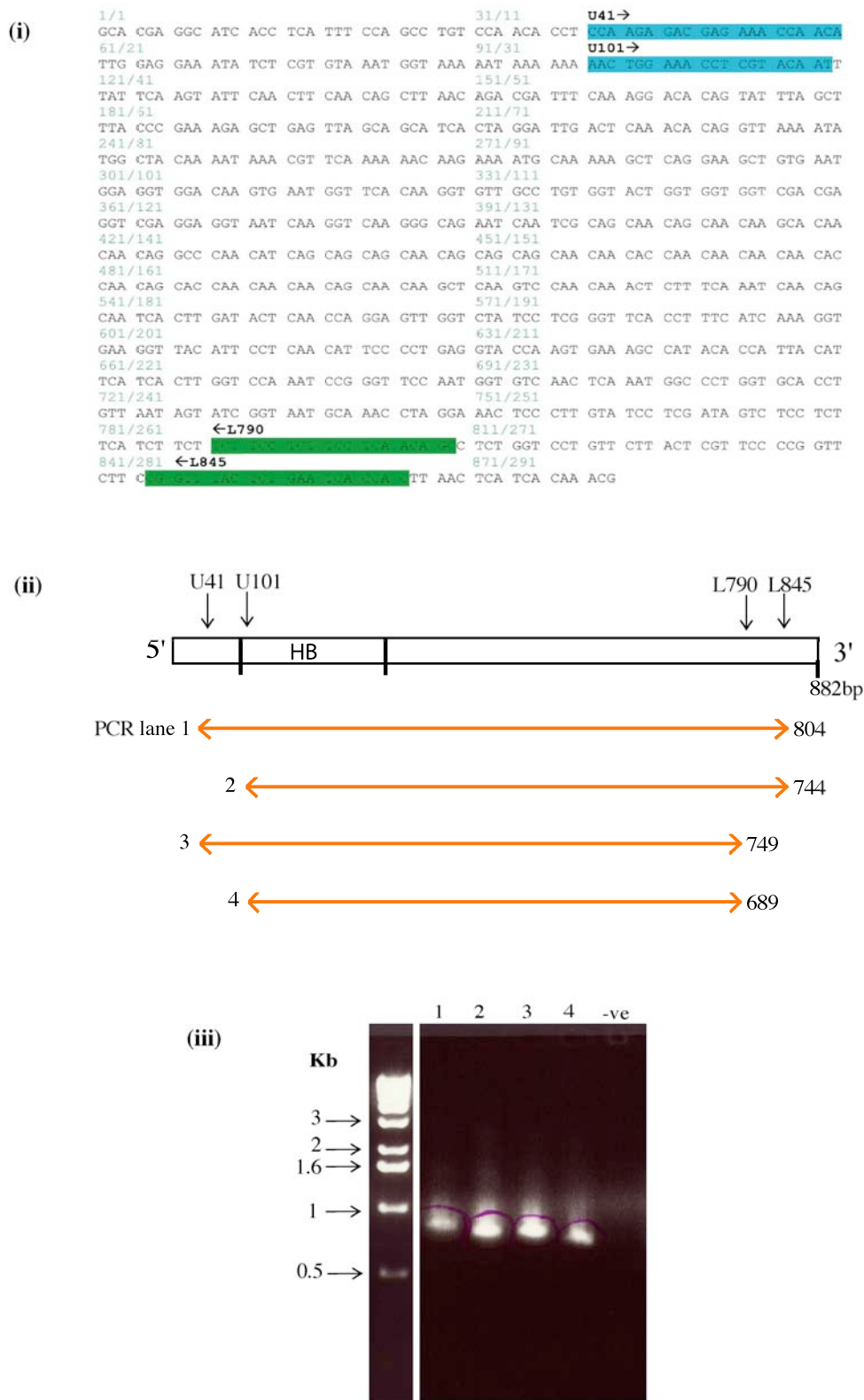


Figure 3.1.4a (i) *Tu-Dll* ORF (882bp) from EST plasmid. Blue and green: Primers used to reamplify from EST fragment in pGEM-T Easy vector. **(ii)** Scaled, schematic representation of EST sequence, with re-amplification primers and resultant PCR fragments illustrated. Fragment sizes predicted from the primer combinations are given, and PCR numbers match lane numbers in **(iii)**, a 1% acrylamide gel of re-amplified *Tu-Dll* DNA fragments (ringed).

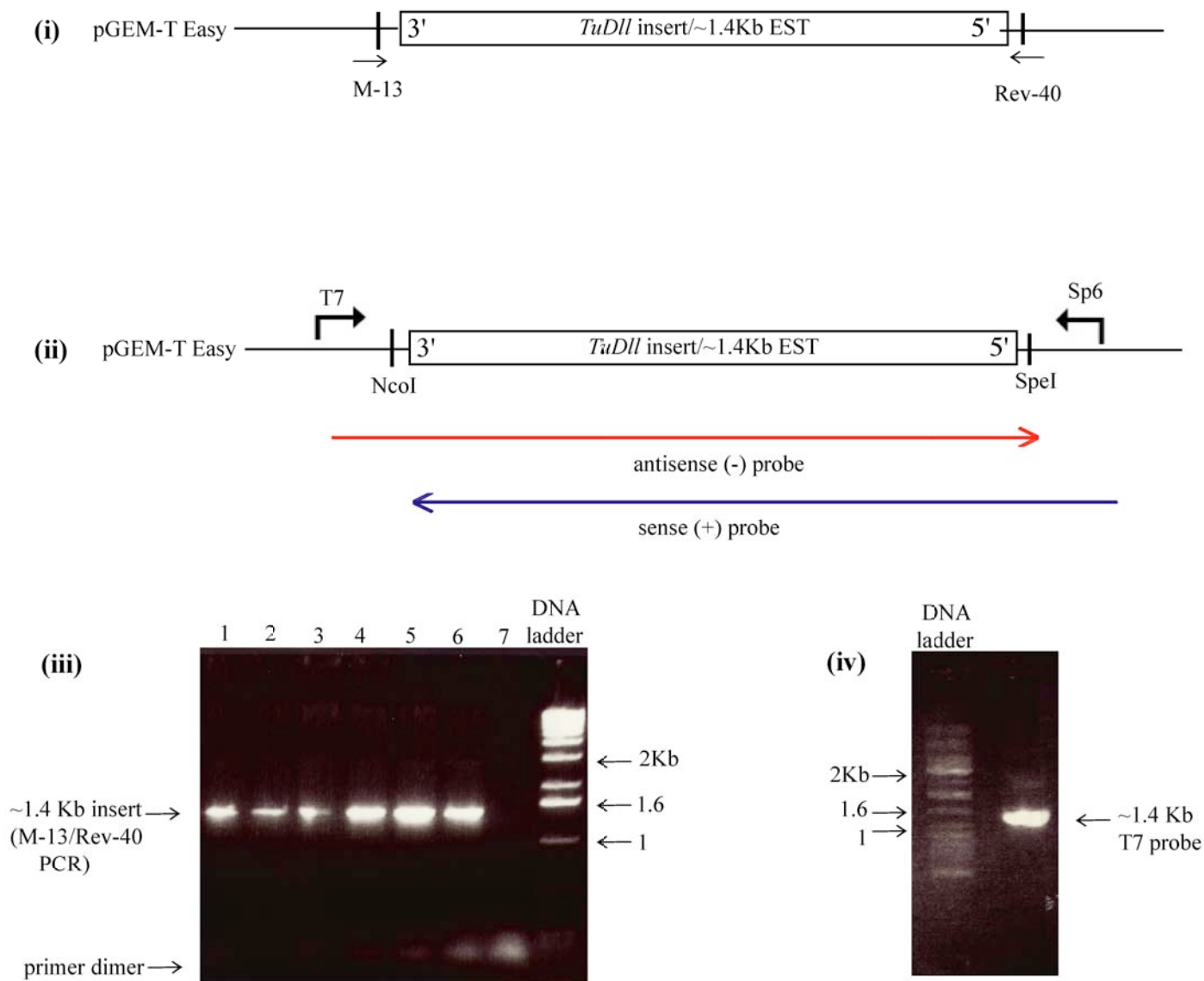


Figure 3.1.4b (i) Schematic illustration of *TuDll* fragment in pGEM-T Easy vector, with forward and reverse primers. (ii) Schematic illustration of mRNA probe synthesis: Correct insert size is confirmed by EcoRI restriction at both ends of the insert or by PCR as in the 1% agarose gel shown in (iii): lane 7 = -ve control. Plasmid is linearised by digestion with either SpeI or NcoI enzyme and ssRNA transcribed by T7 or Sp6 polymerase. T7 transcription on an SpeI-linearised plasmid generates an antisense(-) RNA strand (red arrow), and Sp6 transcription is initiated after linearisation with NcoI to form a sense(+) RNA probe (blue arrow). (iv) Gel confirming efficient SpeI digestion and ssRNA transcription for a *Dll*(-) probe synthesised from EST plasmid start material.

enzymes were deployed for plasmid linearization, restriction and RNA transcription (Figure 3.1.4b,ii).

Probes detected *TuDll* mRNA transcription throughout embryogenesis, although those synthesised from EST plasmid material generally produced the most favourable signal:noise ratio. The shorter probes produced higher background signal that could not be noticeably reduced, as it could for the strongest EST-derived probes, by being applied at 50x standard dilution and allowing a longer AP-H₂O₂ reaction to proceed at lower temperature - 4°C rather than RT (c.f. Materials and Methods, Chapter VIII, section 8.4). Probe hydrolysis had negligible effects on probe efficacy, and dehydration of embryos failed to significantly reduce background staining when this was problematic, and hence neither procedure was routinely used.

3.2 *Tu-Dll* gene expression during embryogenesis

3.2.1 Absence of *Tu-Dll* transcriptional activity, 0-20hr AEL

i) Blastoderm stage (0-9hr AEL)

No *Tu-Dll* mRNA expression was detected by whole-mount *in situ* hybridisation of blastoderm stage embryos (Figure 3.2a,i). Roles for Dll in early development have been proposed for molluscs, arthropods, echinoderms, cephalochordates and vertebrates, but absence of detectable *Dll* mRNA during blastoderm formation fails to support any such role in *Tetranynchus* (Caracciolo et al. 2000; Cohen 1990; Lee and Jacobs 1999; Lowe and Wray 1997). Any possible or hypothetical maternal transcripts could not be detected as early blastoderm formation was the first stage amenable to *in situ* hybridisation, hence ruling out comparison with animals such as mollusc *Mopalia mucosa*, vertebrate *Xenopus laevis* and cephalochordate *Ciona intestinalis* in which maternal *Dlx* transcripts play an active role in establishment of embryonic polarity (Asano et al. 1992; Caracciolo et al. 2000; Lee et al. 2001; Lee and Jacobs 1999).

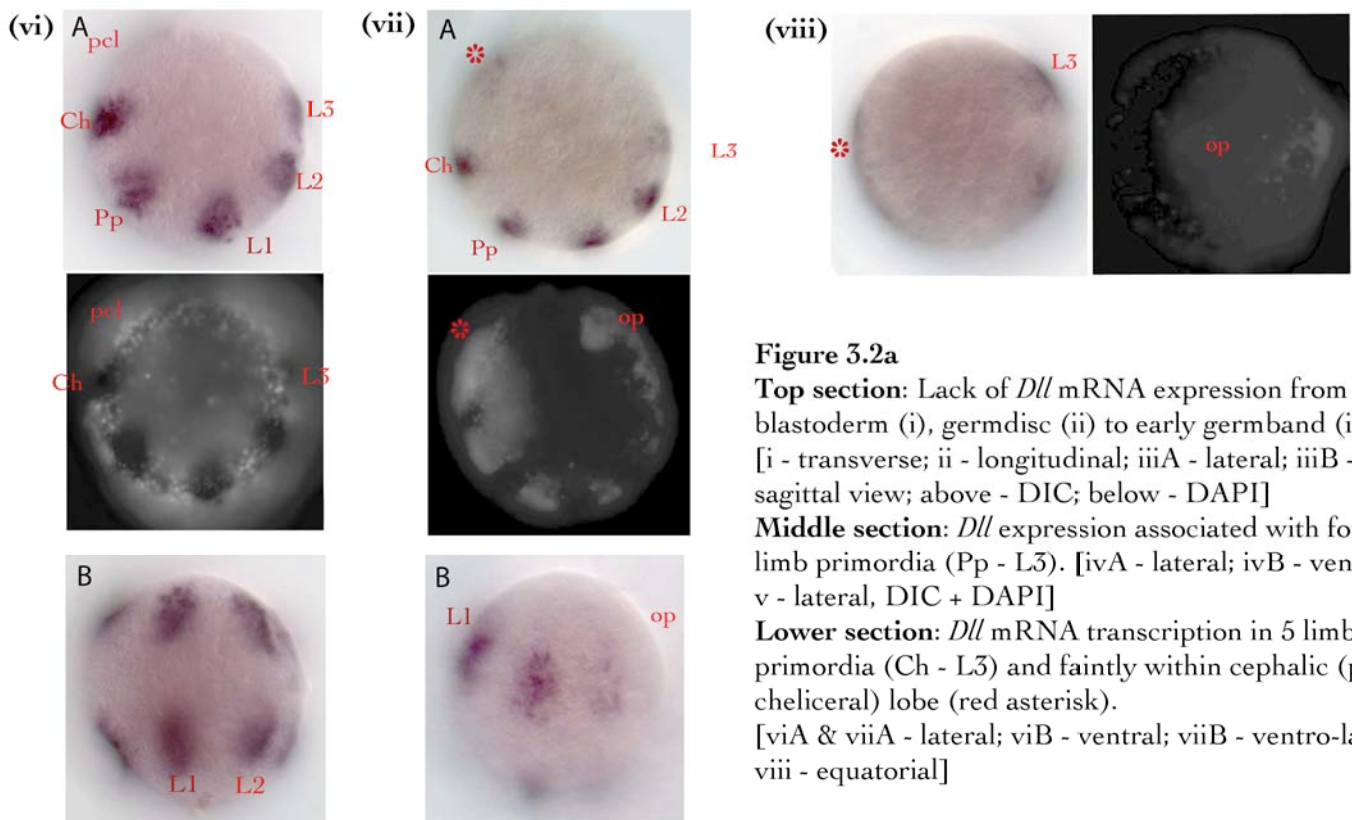
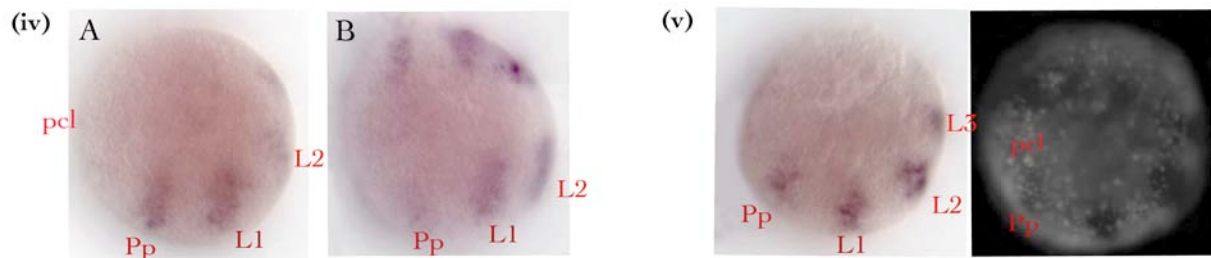
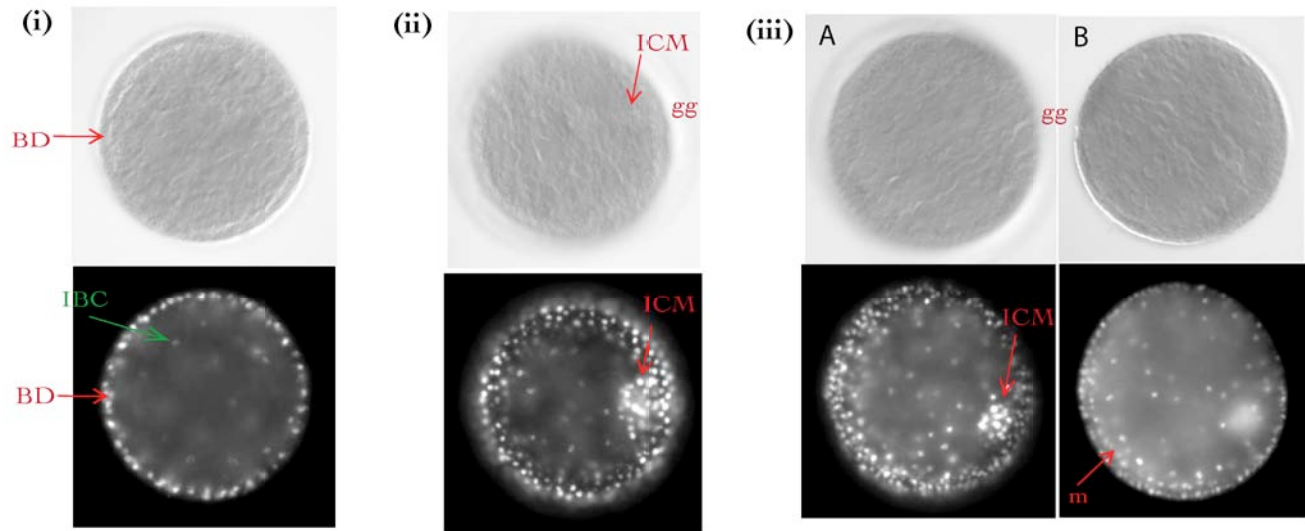


Figure 3.2a

Top section: Lack of *Dll* mRNA expression from blastoderm (i), germdisc (ii) to early germband (iii). [i - transverse; ii - longitudinal; iiiA - lateral; iiiB - sagittal view; above - DIC; below - DAPI]

Middle section: *Dll* expression associated with four limb primordia (Pp - L3). [ivA - lateral; ivB - ventral; v - lateral, DIC + DAPI]

Lower section: *Dll* mRNA transcription in 5 limb primordia (Ch - L3) and faintly within cephalic (pre-chelical) lobe (red asterisk). [viA & viiA - lateral; viB - ventral; viiB - ventro-lat; viii - equatorial]

Abbreviations: pcl - pre-chelical lobe; Ch - chelical segment; Pp - pedipalp; L1 - 1st walking limb segment; op - opisthosomal region; m - mesoderm; ICM: inner/endodermal cell mass; IBC: intra-blastoderm cell; gg: gastrula groove; BD: blastoderm layer.

ii) Germdisc stage (10-15hr AEL) to early germband (16-19hr AEL)

Neither early nor late germdisc stage embryos exhibit evidence for *Dll* mRNA transcription (Figure 3.2a,ii). The earliest germband stage, in which ingressed mesodermal cells are just beginning to migrate anterior-ward, does not show *Dll* expression. This evident transcriptional inactivity continues as the germband extends longitudinally in an anterior direction (Figure 3.2a,iii).

Previous work in molluscs and chordates suggest roles for *Dll* in germ layer formation, ectoderm specification and epithelial development, but lack of *Tu-Dll* activity at germdisc or initial germband stages, in which germ layers are emerging, suggests that such roles are not conserved in chelicerates, or at least not in the Acarida (Caracciolo et al. 2000; Lee and Jacobs 1999).

3.2.2 *Tu-Dll* transcription associated with ventral limb primordia in the late germband (21-22hr AEL)

i) Four paired, limb primordial *Dll* domains

The first sure sign of *Dll* transcription occurs during late germband stage (21hrs AEL), with the simultaneous (or near simultaneous) appearance of four paired patches of mRNA expression each side of the ventral midline. The marked cellular fields denote appendage primordia, corresponding to pedipalps (Pp) and 1st - 3rd walking limb pairs (L1-L3) from anterior to posterior (Figure 3.2a,iv-v). Each domain of *Dll* expressing cells is oval shaped and elongated laterally across 27-30% the width of the germband. Based on indications of overt segmentation, primordia appear in the posterior of each segment; this is typical throughout the arthropods, where segment-polarity gene activity as discussed in Chapter IV is implicated to have a conserved role in para-segmentation and perhaps by association, in directing limb specification.

ii) Five paired, limb primordial *Dll* domains

A fifth, anterior pair of oval *Dll* domains appears just after the activation of *Dll* transcription in Pp-L3 segments (Figure 3.2a,vi-viii). The onset of chelicerate *Dll* expression in the posterior chelicerate segment signifies germband completion, with a full complement of 5 larval appendage-bearing segments; chelicerae, pedipalps and three pairs of walking limbs (Anderson 1973; Boudreaux 1963). A slight delay in chelicerate development relative to other prosomal limbs has been widely reported throughout the Chelicerata, supported by morphological and molecular markers (Anderson 1973; Damen 2002; Simonnet 2005). The feature persists from basal chelicerate groups such as Xiphosura

and Scorpiones to more derived Araneae and Acarida (e.g. *Hyalomma dromedarii* and *Tetranychus urticae*), indicating conservation of certain genetic elements during anterior prosomal segmentation and patterning (Anderson 1973; Simonnet 2005).

Down-regulation in L3

In some embryos *Dll* transcription in L3 primordia appears lower than in segments Ch – L2 (e.g. Figure 3.2a,vii). It is unclear why this should be so: speculatively, genetic interactions repressing overt L4 limb development until later stages may perhaps affect the adjacent L3 segment, transiently reducing *Dll* transcription. Alternatively, the weaker L3 expression may be a mere artefact, derived from random differential staining or perhaps slower relative growth in L3, less mitosis producing lower net mRNA copy numbers for the *Dll* gene.

Limb specification vs. segmentation pattern

The pattern of limb primordium formation in the Sarcoptiform ticks *Hyalomma* and *Ornithodoros* is broadly comparable to that revealed by *Dll* expression in *Tetranychus*, although a clearer posterior to anterior sequential appearance of primordia is observed in the tick species (Anderson 1973). However, the sequence of segment morphogenesis appears to differ in *Hyalomma*, *Ornithodoros* and *Tetranychus*, especially with respect to the mode and timing of opisthosomal formation: in *Ornithodoros* and *Tetranychus* opisthosomal segments are delineated within a pre-existing posterior region of the extended germband, and prosomal limb segments appear to be delineated in a general P-A sequence, walking limbs followed by Pp and Ch segments. By contrast, in *Hyalomma*, five Op segments are added in A-P progression from a growth zone behind the early-formed L4-L3 segments, remaining prosomal segments added *de novo* in P-A progression afterwards (Anderson 1973). This discrepancy among acarids, between broadly conserved prosomal segmentation and divergent opisthosomal segmentation patterns, suggests distinct genetic mechanisms underlying segmentation of these two major tagma. Such an idea supports previous proposals of multiple segmentation mechanisms (e.g. cyclical, Notch-Hairy type vs. pair-rule cascade-type) as being responsible for segment polarity gene activation and formation of anterior and posterior tagma in a range of arthropod groups, from chelicerates to higher insects (Damen 2002; Damen et al. 2005; Damen et al. 2000; Grbic 2000; Patel 1994c; Peel 2004; Peel and Akam 2003; Peel et al. 2005; Schoppmeier and Damen 2005).

iii) Early *Dll* transcription related to neurogenesis

A low level of diffuse *Dll* transcription is evident in a presumptive anterior CNS domain (Figure 3.2a,vii-viii). *Dll* activity in the anterior CNS region in *Tetranychus* is consistent with its observation during neurogenesis in many other metazoans, on the basis of which an ancestral and conserved role

in nervous system development has been proposed (Panganiban 2000; Panganiban and Rubenstein 2002; Prpic and Tautz 2003).

Dll and the arthropod anterior CNS

Distinct ocular lobes or derived (labral) structures are not clear at this early stage of spider mite embryogenesis, but comparison with other arthropods, including *Limulus*, spiders (*Cupiennius*, *Achaearanea*), myriapods (*Glomeris*), crustaceans and insects, suggests that anterior *Dll* expression in this pre-morphogenetic phase in *Tetranychus* is likely to be related to ocular and protocerebral specification (Duman-Scheel and Patel 1999; Mittmann and Scholtz 2003; Panganiban et al. 1995; Popadic et al. 1998; Prpic and Tautz 2003; Schoppmeier and Damen 2001; Younossi-Hartenstein et al. 1997).

Dll and the chordate anterior CNS

Wider phylogenetic comparison can also be made: for example, *Dlx* is active in cephalochordate CNS development, and in vertebrates *Dlx* transcription factors are implicated in forebrain development, specifically in gamma-aminobutyric acid (GABA)-ergic neurons (Brox et al. 2003; Cobos et al. 2005; Panganiban 2000). Deep homology regarding *Dll/Dlx* genes and anterior CNS specification is difficult to prove, as the arthropod protocerebrum and vertebrate forebrain are not traditionally considered homologous; however, conserved gene expression (*otd/Otx+ems/Emx*, *unpg/Gbx2+Pax2/5/8* and *Hox*) in the anterior, mid- and posterior embryonic brain of insects and vertebrates provides evidence to the contrary (Hirth et al. 2003; Sprecher and Reichert 2003; Younossi-Hartenstein et al. 1997). Whether or not insect and vertebrate brains are directly homologous or functionally analogous structures, elements of gene regulation mediated by *Dll/Dlx* could have been conserved in the development of anterior CNS structures in the two disparate groups, representing underlying positional genetic homology, independent of final morphological outcome (Arthur et al. 1999; Hirth et al. 2003; Panganiban 2000; Panganiban et al. 1997; Panganiban and Rubenstein 2002).

3.2.3 *Tu-Dll* transcription in developing limb buds and pre-chelicerar CNS derivatives (23-30hr AEL)

i) Early limb bud stage (23-24hrs AEL)

Once specified, limb primordia grow to form broad, rounded appendage buds. *Dll* expression is strong throughout all but the most proximal part (presumptive coxa-trochanter) of the limb buds, and persists in the pre-chelicerar CNS (Figure 3.2b,i-iii). Presumably, *Dll*'s role in early limb

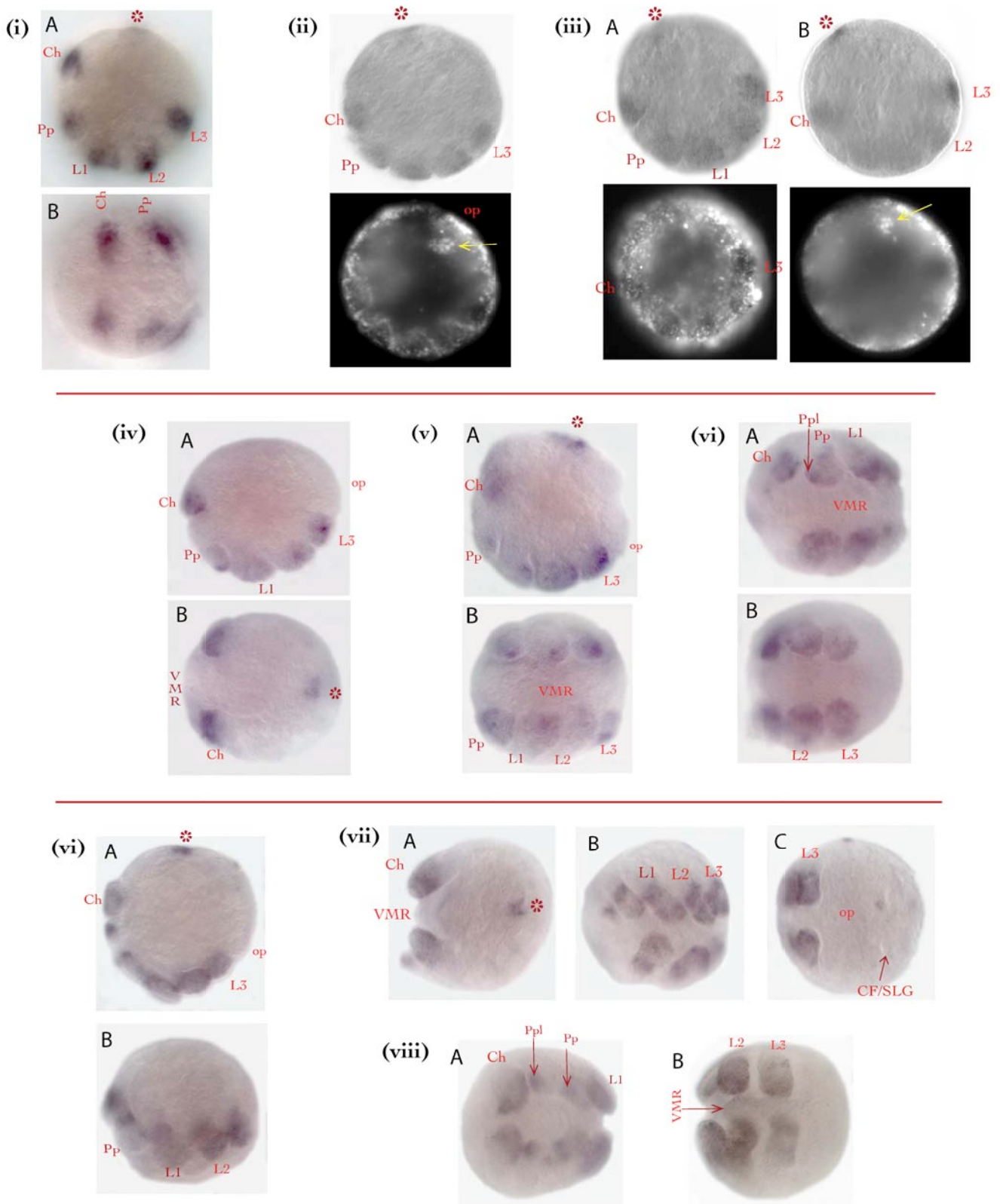


Figure 3.2b Top section: *Dll* expression in early limb buds and pre-chelicerous CNS region (red asterisk). [iA, ii & iiiA - lateral; iB - anteroventral; iiiB - sagittal.] **Middle section:** *Dll* mRNA at late limb bud stage in distal limbs, tips and cephalic lobe/proto-labrum (red asterisk). [iv & vA - lateral; ivB - dorsal; vB & viA - ventral; viB - postero-ventral.] **Lower section:** Ventral ridge contraction stage. *Dll* mRNA in medial reing and distal domains of elongating limbs, and associated with forming labrum structure (red asterisk). [viA - sagittal; viB - lateral; viiA - dorsal; viiB - ventral; viiC - posterior; viiiA - antero-ventral; viiiB - ventral-posterior.]

Abbreviations: Ch - cheliceral segment; Pp - pedipalp; L1 - 1st walking limb segment; op - opisthosomal region; VMR - ventral medial ridge; ICM - inner/endodermal cell mass; CF/SLG - cephalic furrow/semi-lunar groove. Anterior left.

specification has been succeeded by its widely conserved role in proximo-distal appendage patterning, assigning distal limb, or telopod, identity (Boxshall 2004; Haas et al. 2001a; Haas et al. 2001b; Panganiban 2000; Schoppmeier and Damen 2001).

ii) Late limb bud stage (25-30hrs AEL)

P-D patterning

As development progresses, appendages lengthen and *Dll* transcription is detectable throughout distal limb segments (Figure 3.2b,iv-vi). Restriction of *Dll* activity to the distal limb (presumptive femur/genu-tarsus) is a key step in proximo-distal axis patterning in many arthropods, and I shall assume that this pattern, with many or all of its associated genetic controls, has also been conserved in *Tetranychus* (Boxshall 2004; Panganiban 2000; Schoppmeier and Damen 2001).

PNS sensory organ specification

In some embryos *Dll* seems exceptionally up-regulated at appendage tips, possibly related to PNS specification in distal tarsi (Figure 3.2b,ivA,v-vi). In the case of *Tetranychus*, this would involve specification of tarsal empodia that bear sensory setae of various kinds and, in palp tarsi, solenidia modified for silk secretion (Návia, 2004; Boxshall, 2004; Krantz, 1979; Grandjean, 1935; Grandjean, 1948; Boudreaux, 1963; RBINS, 2006). *Tetranychus* chelicerae become highly modified to form stylets, so secondary distal *Dll* upregulation may serve to control genetic networks for limb-specific morphogenesis and sensory organ formation, and/or provide a separate identity from the proximal cheliceral segments, which form the stylophore body (for more details on gnathosomal development, see Chapter II).

Early labrum formation

Compared to broader previous *Dll* expression, the ‘anterior CNS’ domain constricts to a small, bilaterally symmetrical but medial domain (Figure 3.2b,ivB-vA). This may reflect a process evident in myriapods, spiders, mites and insects, wherein *Dll* is initially expressed in ocular lobes, but later becomes stronger in the labrum as it forms at the midline (Panganiban et al. 1994; Popadic et al. 1998; Prpic and Tautz 2003; Schoppmeier and Damen 2001; Thomas and Telford 1999). The restricted domain in *Tetranychus* would then correspond to an early labrum, forming either independently, or from rearrangement of, ocular lobe-derived cells. Bilateral clusters of *Dll*-positive cells appear to move from ocular lobes to the labrum in *Archegozetes longesitosus* (oribatid mite), favouring a segmental appendage origin (Thomas and Telford 1999). Although I could not observe patterns of anterior cell migration in *Tetranychus*, the bilobed form of the emerging labrum (Figure 3.2b,iv&vii) is suggestive of a derived limb affinity: this is discussed further in section 3.2.4(ii).

3.2.4 Dynamic *Tu-Dll* transcription during ventral ridge contraction (30-39hr AEL)

i) Modulated *Tu-Dll* activity along the P-D limb axis

Proximo-distal 'ring-and-tarsus' Dll domains in Tetranychus

During ventral ridge contraction, pedipalps and walking limbs elongate and *Dll* transcription becomes down-regulated between the distal and proximal parts of the earlier telopod domain, generating a medial ring-plus-tarsal-tip expression pattern (Figure 3.2b,viB,viiB,viiiA). Compared to patterns of *Dll* expression along the P-D limb axis in other arthropods (summarised in the following paragraph), the spider mite Pp-L4 'ring and tarsus' pattern resembles most closely that seen in higher insects. The large phylogenetic disparity between the chelicerate and hexapod lineages suggests that this uniquely shared similarity may be due to convergent evolution, or parallel variation of an ancestral arthropod limb patterning mechanism, each domain along the P/D axis related to taxon-specific differences such as leg length, joint formation, and sensory organ innervation.

Proximo-distal Dll domains in other arthropods

Tetranychus 'ring and tarsus' *Dll* expression pattern is reminiscent of the typical 'ring and sock' domains seen in holometabolous insect thoracic limbs, including the dipteran *Drosophila* (at leg disc eversion, ~5hr after pupariation) and lepidopteran *Precis* (from ~30% embryo development): increased *Dll* domain complexity is presumed to reflect new roles in P-D limb patterning, compared to earlier deployment in limb precursor cell specification (Diaz-Benjumea et al. 1994; Panganiban et al. 1994). In the hemimetabolous orthopteran *Achaeta* and basal hexapod *Thermobia*, a 'sock' domain is accompanied by a double ring, resolving either from or instead of the single ring observed in *Drosophila* and *Precis* (Abzhanov and Kaufman 2000c; Popadic et al. 1998). Beyond Insecta the 'ring(s) and sock' pattern is not clearly conserved: myriapods generally retain uniform *Dll* expression in all telopod segments but the trochanter; crustaceans and the mite *Archegozetes* apparently retain the same, but spiders diverge from previously seen patterns. In the pedipalp and L1-L4 limbs of *Achaearanea*, *Cupiennius* and *Steatoda* (Araneae), *Dll* is expressed irregularly along the P-D axis, generating a multiple-ring pattern (Abzhanov and Kaufman 1999; Abzhanov and Kaufman 2000c; Panganiban et al. 1995; Popadic et al. 1998; Prpic and Tautz 2003; Thomas and Telford 1999). Minimal available *Dll* antibody staining data for *Limulus polyphemus* (Chelicerata: Xiphosura) indicates translational modulation during appendage and sense organ development, expression in limbs changing from distal to medio-distal and later to distal tips, with lower levels of expression in the telopod (Mittmann and Scholtz 2001).

Cheliceral and pedipalp lobe Dll domains in Tetranychus

Chelicerae and pedipalp lobes in *Tetranychus* both retain uniform distal *Dll* expression during ventral ridge contraction (Figure 3.2b,viiA,viiiA). The pattern matches that observed in the mite *Archegozetes*, and in the spiders *Achaearanea*, *Cupiennius* and *Steatoda* where *Dll* activity is uniform in the cheliceral telopod (and later opisthosomal dorsal limb rudiments, e.g. spinnerets, tubular tracheae) (Popadic et al. 1998; Schoppmeier and Damen 2001; Thomas and Telford 1999). This comparison suggests conserved *Dll* regulation and function in pre-morphogenetic stages of gnathosomal development within Acarida and Araneae.

ii) *Dll* activity associated with the labrum

At the anterior midline of the pre-cheliceral lobe, the labral ectoderm continues to transcribe *Dll*. The nascent labrum is still bilobed, with anterior lobes delineated most clearly, and it migrates posteriorly, sustaining a position anterior to the stomodeum (Figure 3.2b,viA,viiA).

The affinity of the arthropod labrum?

The segmental or non-segmental affinity of the labrum has been a topic of long-standing debate, yet to be resolved (Haas et al. 2001a; Haas et al. 2001b; Rempel 1975). The Euarthropod labrum is commonly bilobed, and correlated with *Distal-less* expression - and hence appendage identity - in species as phylogenetically disparate as *Limulus polyphemus* (Xiphosura) and *Drosophila melanogaster* (Diptera) (Mittmann and Scholtz 2001; Panganiban et al. 1994; Panganiban et al. 1995; Prpic and Tautz 2003; Scholz et al. 1997; Thomas and Telford 1999). However, in contrast to *Dll* specifying only appendage identity, certain analyses of arthropod *Dll* activity have linked its presence in the elusive labrum, as well as in posterior structures such as anal valves, circae and the telson, to a role as a genetic marker for 'extreme body axis regions'. (Panganiban et al. 1995; Popadic et al. 1998; Popadic et al. 1996; Scholtz 1997). The possibility that *Dll* marks segmental appendages as well as non-appendicular, or non-segmental, terminal derivatives, means that other morphological or genetic means of assessing labrum affinity are also desirable:

An appendage identity has been hinted at in the isopod crustacean *Armadillidium* and strongly indicated by a homeotic mutation in the flour beetle *Tribolium castaneum*: Incipient distal fusion of two limb-like lobes occurs during labrum ontogeny in *Armadillidium*, and the *Tribolium* homeotic mutation *Antennaglea-5* (*Ag*⁵) transforms the labrum into gnathal-type limbs of mixed mandibular and maxillary identity, medial and lateral plates resembling gnathobasal endites (Haas et al. 2001a; Popadic et al. 1998). The *Ag*⁵ phenotype implies labrum origin from a fused pair of endites (i.e. coxopod-derived), but this may not be true for all arthropods: in the spider *Cupiennius*, *Dll* RNAi removes all telopod-derived structures, including the whole labrum which is therefore designated some form of telopod

affinity (Schoppmeier and Damen 2001). *Wingless* and *engrailed* segment polarity gene expression, and several histological features (e.g. distribution of coelomic sacs, mesoderm, cerebral innervation) are variable among the arthropods thus far studied, confounding assignment of the labrum to any potentially novel anterior or extant segment (Damen 2002; Damen et al. 1998; Mittmann and Scholtz 2003; Popadic et al. 1998; Scholz et al. 1997; Telford and Thomas 1998a).

The affinity of the spider mite labrum?

Tetranychus provides further support for the appendicular nature of the labrum: a bilateral distribution of *Dll*-positive cells during its formation hints that, as in *Archegozetes longesitosus*, the labrum may develop by fusion of a pair of much altered, or vestigial, appendages (Thomas and Telford 1999). In addition, *Dll* transcription correlates with telopod identity in other prosomal appendages (Ch – L3), so a similar identity (albeit modified) could be inferred for the labrum; functional studies are essential to confirm or refute this possibility.

3.2.5 *Tu-Dll* transcriptional modulation within appendages during ventral closure (40-45hr AEL)

i) *Dll* modulation in anterior prosomal limbs and the labrum

Completion of ventral closure is accompanied by rotation of chelicerae and pedipalp lobes over the stomodeum, bringing them into apposition prior to further gnathosomal morphogenesis. *Dll* transcription is restricted to the most distal margin of the chelicerae and pedipalp lobes, correlated with the point at which rotated, paired limbs approach at the midline (Figure 3.2c,iA,iiiB). At this stage, *Dll* mRNA is detected in the labrum as an apparently double, medial stripe (Figure 3.2c,iiiA); this pattern may reflect two adjacent, albeit derived distal limb elements, supporting the appendicular affinity of the labrum and typical conservation of distal *Dll* expression.

ii) Persistent *Dll* activity in ‘ring and tarsus’ tip domains

Dll continues to be expressed in ‘ring and tarsus’ domains along the proximo-distal axis during elongation and overt segmentation of the pedipalps – L3 limbs (Figure 3.2c,i-iii). The ends of the limbs develop a subtle biramous form relative to earlier stages, correlated with restriction of tarsal *Dll* expression to a more focussed domain between the two developing distal lobes. I hypothesise that *Dll* modulation in these limbs may be related to control of allometric differentiation, development of joints, and - particularly at the distal tip - sensory organogenesis.

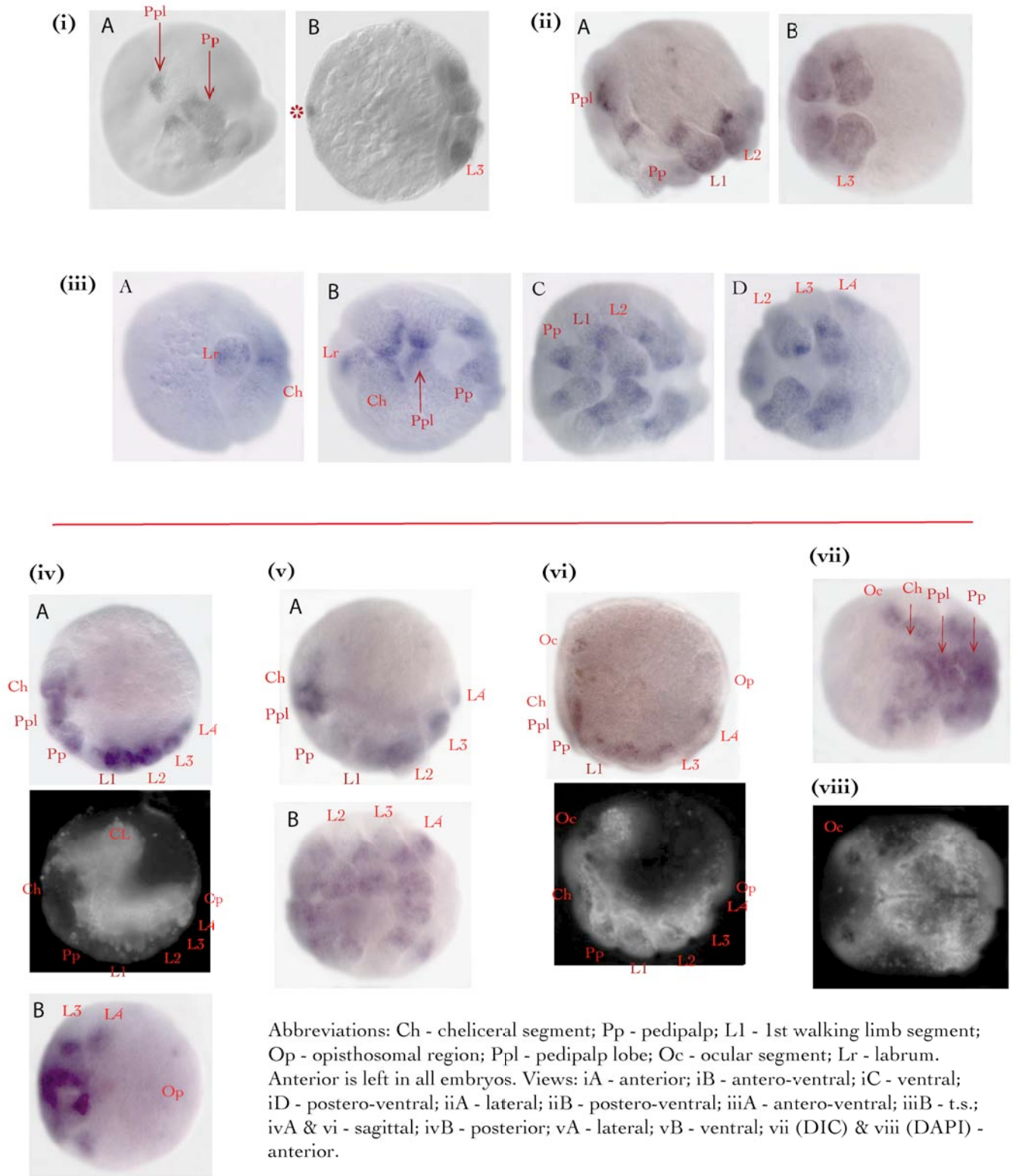


Figure 3.2c *Dll* whole mount *in situ* mRNA expression during ventral ridge and dorsal closure.

Top section: Ventral closure and limb elongation stage. *Dll* expression in all appendages including labrum and emerging L4 limb bud. Pattern in distal and medial ring domains along elongating P-D Pp-L3 axes. Lower section: Germband contraction (to ventral surface) and dorsal closure stage. *Dll* mRNA detected in distal and medial ring domains, as well as in the L4 limb buds and associated with the ocular lobes/anterior CNS. No transcription in the very reduced opisthosoma.

iii) *Dll* activation in emerging L4 primordia

In late ventral closure stage embryos, primordial L4 limb buds appear as small swellings at the prosomal-opisthosomal (disjugal) boundary, each marked by a comparatively low level of *Dll* transcription (Figure 3.2c,iiiD). As a molecular marker for limb ectoderm, *Dll* mRNA in L4 cells supports speculations in this and other acarids, that L4 limbs are specified in late embryogenesis but that outgrowth and patterning are delayed until the protochrysalis stage (Anderson 1973; Boudreaux 1963; Krantz and Lindquist 1979). The small size of the swelling and low level of *Dll* transcription indicate recent activation, placing genetic specification of these otherwise repressed limbs between 42-45hrs AEL, i.e. approximately 20hours after other prosomal limbs were specified.

3.2.6 *Dll* gene activity in appendages and the ocular segment during germband contraction and dorsal closure (45-50hrs AEL)

i) Prosomal appendage *Dll* domains

Dll is expressed throughout Ch – L3 limbs as previously, at the distal tip and in a ring domain positioned approximately half way along the proximo-distal limb axis (Figure 3.2c,iv-v). Persistence of these domains indicates a sustained requirement for *Dll*, hypothetically related to late PNS and P-D axis patterning functions. Higher levels of uniform *Dll* transcription are observed in the L4 primordia, associated directly or indirectly with their growth to form clear buds behind L3 leg bases (Figure 3.2c,iv-v).

The labrum is a transient structure, no longer visible morphologically or as *Dll*-positive cells at this late stage of embryogenesis (Figure 3.2c,iv-viii). According to Anderson (1973) the acarid labrum fades away during cheliceral rotation over the stomodeum, but more recent morphological studies of acarids suggest that it can be either lost during ontogeny (e.g. *Tetranychus gigas*, *T. bunda*, *Erynetes*, *Tydeus*, *Hyalomma*, *Ornithodoros*) or remain to constitute a reduced but functional mouthpart (e.g. *Archegozetes longesitosus*) (Thomas, 1999; Anderson, 1973; André, 2003; Flechtmann, 2002).

ii) Ocular segment *Dll* domains

Dll mRNA is detectable in developing ocular segments (Figure 3.2c,vi-viii), involved in directing protocerebral segment development and hypothetically also affecting ontogenetic pathways in other

anterior CNS components, as shown in *Limulus*, spiders and hexapods (Mittmann and Scholtz 2001; Prpic and Tautz 2003).

3.3 *Tetranychus* Distal-less protein: antibody tests

The polyclonal Dll antibody

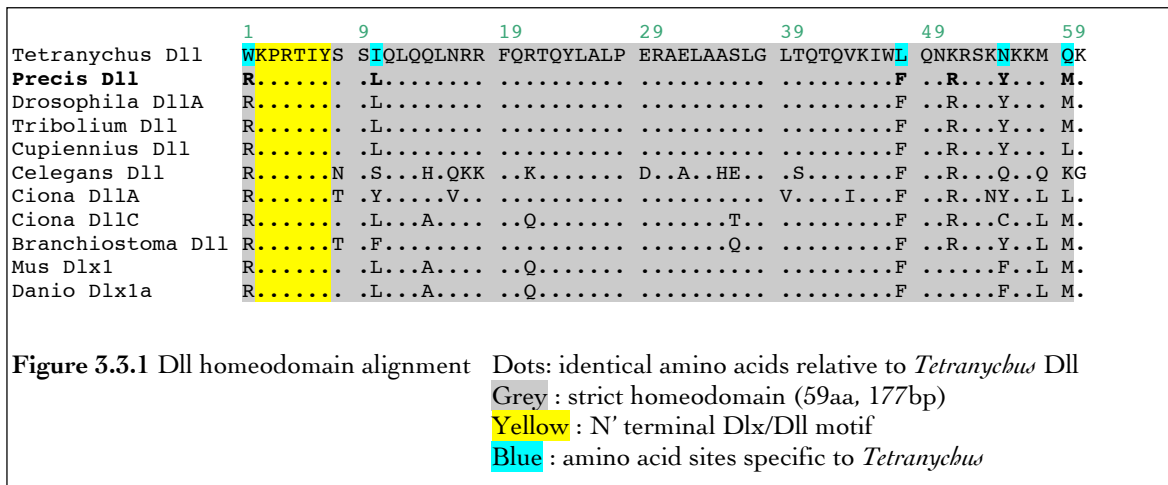
Panganiban et al. (1995) produced a polyclonal Dll antibody, raised against part of *Precis coenia* (Lepidoptera) Dll protein - the homeodomain and a few amino acids either side. This antibody was shown to recognise Dll and Dlx proteins in appendages of a range of arthropods as well as in polychete parapodia, onychophoran lobopodia, echinoderm tube feet, ascidian ampullae and vertebrate limb buds (Panganiban et al. 1994; Panganiban et al. 1995).

Lack of antibody reactivity in *Tetranychus* embryos

A sample of the Panganiban α -Dll antibody was available to me, and I tested it for potential to detect Dll protein expression in whole-mount *Tetranychus* embryos of mixed stages. I sought to detect α -Dll reactivity at a range of primary antibody dilutions (1:10, 1:100 and 1:500), but did not observe target protein expression at any developmental stage. This negative result suggests that either:

- i) the specific epitope bound by the polyclonal α -Dll is divergent in *Tetranychus* relative to the original *Precis* protein, preventing cross-reactivity, or
- ii) the antibody had degraded (it was several years old) to a level at which it no longer effectively binds Dll/Dlx proteins *in situ*, or
- iii) Dll protein structure was denatured or disrupted during spider mite embryo fixation, preventing effective binding or recognition of the antibody to its epitope.

To address above options for failed antibody staining, I first constructed an alignment of the *Tetranychus* Dll homeodomain with the *Precis coenia* Dll homeodomain (i.e. original antigen sequence), including Dll orthologs in several species that positively cross-react with the polyclonal α -Dll antibody (see Figure 3.3.1):



Two of the five *Tetranychus*-specific Dll residues (trp/w + leu/L) correlate with consistently alternative residues in all other species, each amino acid notably 100% conserved among these cross-reactive species. These conserved amino acids may represent critical epitope binding sites that are missing in *Tetranychus* Dll, but their physical separation, lack of correlation with other scattered divergent residues (Figure 3.3.1) and lack of functional evidence makes this possibility only tentative. Furthermore, Dearden et al. (2002) report positive detection of Dll protein with the Panganiban antibody in a Canadian population of *Tetranychus urticae*, so I assume that lack of cross-reactivity that I obtained with the same antibody was merely apparent, and not caused by epitope disparity. Rather, options (ii) or (iii) must explain the negative result: degradation either of the antibody over time, due to freeze-thawing, or excessive degradation of Dll protein during embryo fixation. The fixation protocol minimises steps that could destabilise protein structure (see Chapter VIII: Materials and Methods, sections 8.1.4 and 8.5), although it is different than the Dearden et al. (2002) protocol (Dearden et al. 2002).

3.4 Cloning *Tetranychus Sp8/9* and *Sp1/3/4* orthologs

3.4.1 Degenerate PCR screening

i) Degenerate PCR

Through a degenerate PCR screen of *Tetranychus urticae* genomic DNA, I isolated multiple clones for two different *Sp* genes, provisionally termed *Sp-A* and *Sp-B*, and later shown to be likely orthologs of *Sp8/9* and *Sp1/3/4* genes respectively (Figure 3.4.1a).

Sp family genes all include code for a signature ‘Sp’ domain, ‘Buttonhead’ domain and a triple zinc finger (ZnF) domain with multiple paired cysteine and histidine (‘cys2his2’) residues, critical in zinc finger secondary structure (Bouwman, 2002; Kawakami, 2004; Bell, 2003; Schöck, 1999; Athanikar, 1997). I targeted degenerate primers within the most conserved Btd and ZnF domains (see Appendix 2.2 for primer sequence details), and amplified several DNA fragments of ~180bp as predicted. Although I tested my own primers, PCR was most successful with primers kindly donated by M. Hildebrant, Universität zu Köln: primary touchdown PCR with Sp-F1b + Sp-R1/R2, and nested standard PCR with Sp-F2 and Sp-R2. Touchdown PCR program downdeg46 (see Ch VIII: Materials and Methods, section 8.2) was employed primarily, with highly stringent early cycles to selectively target *Sp* genes, and less stringent later cycles for optimal amplification of *Sp* products. Time constraints unfortunately prevented me from extending these preliminary *Sp* gene fragments, but such extra sequence data are desirable to confirm orthology inferences.

ii) Sequence analysis

A comparison of nucleotide and amino acid data (see Figure 3.4.1a) for the two putative *Tetranychus Sp* genes, *Sp-A* and *Sp-B*, reveals that they share 70.1% nucleotide identity (120/170bp) and 72% identity at amino acid level (41/57 residues).

A multiple alignment of orthologous Sp proteins from a range of arthropod and deuterostome taxa (Figure 3.4.1b), over a region corresponding to nucleotides used in subsequent Bayesian analysis (section 3.4.2), shows full conservation of diagnostic Buttonhead and Zn finger domains, including cys2his2 residues. Within the alignment, a disparity can be seen between the major two branches of the Sp family: Sp6/7/8/9 and Sp1/3/4/5 (Bouwman and Philipsen 2002; Kawakami et al. 2004). *Tu Sp-A* appears to share closest similarity with Sp8/9 orthologs, bearing sub-group specific Btd and ZnF domains residues (Figure 3.4.1b). *Tu Sp-B* contains motifs between the Btd and Zn finger

(i)

```

1/1 31/11
Sp-A GCG ACG TGT GAC TGC CCG AAC TGT CAG GAT GCA GAG AAA CTT GGT CCT GCC
Sp-B GCA ACG TGC GAT TGC CCG AAC TGT CAG GAG GCG GAG GGT AGA AAT --- ---

Sp-A ala thr cys asp cys pro asn cys gln asp ala glu lys leu gly pro ala
Sp-B ala thr cys asp cys pro asn cys gln glu ala glu gly arg asn --- ---

61/21 91/31
GGG ATT CAT GCA AAG AAA AAG CAA ACT CAC AAT TGT CAT ATT CCC GGA TGT GGT AAA GTT
--- --- AGT GAA ACA AAA AAG AAG CAA CAT AAT TGT CAT ATT CCT GGT TGT AAC AAG GTA

gly ile his ala lys lys lys gln thr his asn cys his ile pro gly cys gly lys val
--- --- ser glu thr lys lys lys gln his ile cys his ile pro gly cys asn lys val

121/41 151/51
TAC GGA AAA ACT TCT CAC CTT AAG GCT CAC CTT CGC TGG CAC ACC GGC GAA CGA CCC TT-
TAT GGT AAA ACA TCT CAT CTT CGA GCT CAC CTT CGT TGG CAC ACT GGT GAG CGC CCC TT-

tyr gly lys thr ser his leu lys ala his leu arg trp his thr gly glu arg pro phe
tyr gly lys thr ser his leu arg ala his leu arg trp his thr gly glu arg pro phe
```

(ii)

```

1 11 21
Sp-A (Sp8/9) ATCDPCNCQD AEKLGPAGIH AKKKQTHNCH
Sp-B (Sp1/3/4) ATCDPCNCQE AEGRN-----S ETKKKQHICH

31 41 51
Sp-A (Sp8/9) IPGCGKVYGK TSHLKAHLRW HTGERPF
Sp-B (Sp1/3/4) IPGCNKVYGK TSHLRAHLRW HTGERPF
```

Figure 3.4.1a

- (i) Comparative nucleotide and coding sequence data for *Tetranychus urticae* *Sp-A* (*Sp8/9* ortholog) and *Sp-B* (*Sp1/3/4* ortholog). Underlined: target sequence of degenerate primers MHfw1b and MHrev1, 5' and 3' respectively. **Yellow**: Distinct bases (50 of 170 = 29%) and amino acid residues (16 of 57 = 28%); 72% conservation. **Red** text: Btd domain (minus N' gly + arg). **Blue**: Zn finger domain, incomplete.
- (ii) Comparative polypeptide alignment for *Tetranychus* *Sp-A* (*Tu-Sp8/9*) and *Sp-B* (*Tu-Sp1/3/4*) proteins. **Yellow**: divergent amino acid sites (28%). **Red**: Btd domain. **Blue**: ZnF domain (partial).

Btd domain													Zn finger domain												
Dm_Sp8	RYAGRATDC	PNCQEAERLG	HLRKNHSC	HIPGCGKVG	KTSHLKA	LR	WTGERPFCNWLFC	GKRFT	RSDELOR	LR	TTTGKRFAC	PVCNKRFRMG	D	LAKAVKTH											
Tu_SpA	---	...	D..K..	AK..QT.N	---	---	---	---	---	---	---	---	---	---											
Ag_Splk	..S											
Apis_Splk	..T	I											
Tc_Splk	..T											
Mus_Sp9	..S											
Danio_Sp9	..S											
Sp_Sp9	..S											
Homo_Sp8a	..S											
Homo_Sp8b	..S											
Mus_Sp8	..S											
Danio_Sp8	..S											
Mus_Sp7	SG..S											
Homo_Sp7	SG..SS											
DanioSp7	STT..SS											
Tu_SpB	---	...	GRN SET..KQ.I	N	---	---	---	---	---	---	---	---	---	---											
Homo_Sp1	RTR.EA.T	Y.KDS.GR	DPG..KQ.I	Q	---	---	---	---	---	---	---	---	---	---											
Mus_Sp1	RTR.EA.T	Y.KDS.GR	DPG..KQ.I	Q	---	---	---	---	---	---	---	---	---	---											
Danio_Sp1	RTR.EA.T	Y.KDG.GR-	DPS..KQ.I	---	---	---	---	---	---	---	---	---	---	---											
Mus_Sp3b	KRLR.VA.T	K.GGGR	N.G..KQ.I	---	---	---	---	---	---	---	---	---	---	---											
Mus_Sp3a	KRLR.VA.T	K.GGGR	N.G..KQ.I	---	---	---	---	---	---	---	---	---	---	---											
Homo_Sp3	KRLR.VA.T	K.GGGR	N.G..KQ.I	---	---	---	---	---	---	---	---	---	---	---											
Rattus_Sp4	KRLR.VA.S	R.G.GR	EPG..KQ.V	---	---	---	---	---	---	---	---	---	---	---											
Mus_Sp4	KRLR.VA.S	R.G.GR	EPG..KQ.V	---	---	---	---	---	---	---	---	---	---	---											
Homo_Sp4	KRLR.VA.S	R.G.GR	EPG..KQ.I	---	---	---	---	---	---	---	---	---	---	---											
Danio_Sp4	KRLR.VA.S	RDG.GRN	DPS..KQ.V	ME	---	---	---	---	---	---	---	---	---	---											
Mus_Sp5	TARRCRR.R	A.GGAP	EPG..KQ.V	V	---	---	---	---	---	---	---	---	---	---											
Homo_Sp5	TARRCRR.R	A.GGAP	EPG..KQ.V	V	---	---	---	---	---	---	---	---	---	---											
Danio_Sp5	TARRCRR.R	SS--SS	EPG..KQ.I	---	---	---	---	---	---	---	---	---	---	---											
Dm_btd	LQRRSVR.T	TNEMSGL	DE.GRKQ.I	---	---	---	---	---	---	---	---	---	---	---											
Homo_TIEG1	GTRLSPIAPA	GFSPSAAKV	DSSRIRS.I	SH	---	---	---	---	---	---	---	---	---	---											
Homo_TIEG2	NTKLLPLAPA	VFTITSSQNC	DFSRRNYV	SF	---	---	---	---	---	---	---	---	---	---											
Homo_KLF	VT.AG.VKSG	QSDSDQGG	PEN..RV.R	QFN	---	---	---	---	---	---	---	---	---	---											

Figure 3.4.1b Multiple amino acid sequence alignment for Sp proteins used in Bayesian analysis (alignment calculated in PHYMLIP v.3.6 ProtDist using Jones-Taylor-Thornton model parameters). Dots (.): identical amino acid relative to the reference sequence (*Drosophila* Sp8). Gaps (-): absent data. Sp family domains: **red** - Btd domain; **blue** - zinc finger domain. Pairs of cysteine (C) and histidine (H) residues (cys2his2), responsible for zinc finger structure, are highlighted in turquoise and green respectively (Bouwman & Philipson, 2002). Bold: *Tetranychus* Sp-A and Sp-B proteins. By eye, the *Tetranychus* Sp sequences seem related to different genes, possibly belonging to groups Sp8/9 and Sp1/3/4. Abbreviations: Tu = *Tetranychus urticae*, Tc = *Tribolium castaneum*, Dm = *Drosophila melanogaster*, Ag = *Anopheles gambiae*, Sp = *Strongylocentrotus purpuratus*. Homo outgroup sequences: TIEG - TGF-beta early inducible protein, KLF - ubiquitous Krüppel-like factor.

domains (e.g. GRN, KQHI), and a Zn finger arginine that support Sp1/3/4 affinity (Figure 3.4.1b). However, the Sp-B Btd domain itself is more similar to that of Sp8/9 proteins: the Sp1/3/4/5 sequences have fairly consistent Btd box modifications relative to the Sp7/8/9 genes, particularly in scattered 5' residues (Bouwman and Philipsen 2002). The apparently confounding affinities of Sp-B Btd domain vs. more C' and ZnF domains may be due to:

- i) stochastic substitution patterns, or
- ii) selection pressure to conserve Btd box sequences and preserve a specific functionality trait, or
- iii) conversion of Btd boxes (i.e. domain swapping) between an *Sp8/9*-type gene and an *Sp1/3/4* gene, giving the effect of *Sp8/9* affinity; this would be similar to the recurrent homeobox conversion and concerted evolution in hexapod Engrailed genes (Peel et al. 2006).

3.4.2 Phylogenetic analysis of *Sp* genes

i) Full dataset results

To further clarify spider mite *Sp* gene orthologies, I carried out a Bayesian inference analysis of aligned nucleotide data from numerous arthropod and deuterostome *Sp* orthologs, as well as a few outgroup gene sequences (*Homo TGF- β inducible early protein* and *ubiquitous Krüppel-like factor*). The resultant consensus topology (Figure 3.4.2a) identifies *Tu Sp-A* most strongly as an *Sp8/9*-type gene, and *Sp-B* as an *Sp1/3/4* gene. Reliability of the topology, and hence gene orthology assignments, is supported firstly by maximal posterior probability separating outgroup genes from the main *Sp* family clade, and secondly by support for distinct *Sp1/3/4/5* and *Sp7/8/9* clades (Beermann et al. 2004; Kawakami et al. 2004). In addition, within the *Sp* clade itself all sequences except *Drosophila bt δ* have standard branch lengths, potentially reducing the likelihood of topological error derived from long branch effects. The more distantly related *TEIG* and *KLF* genes gave relatively long branches, and their effects on *Sp* clade resolution were not clear.

ii) Reduced dataset results

In order to optimise resolution within the *Sp* clade, I repeated the Bayesian analysis without the long branch 'outgroup' sequences. Further resolution of relationships was recovered: *Tu Sp-A* falls within a monophyletic *Sp7/8/9* clade receiving higher (0.92) probability support than in the larger analysis,

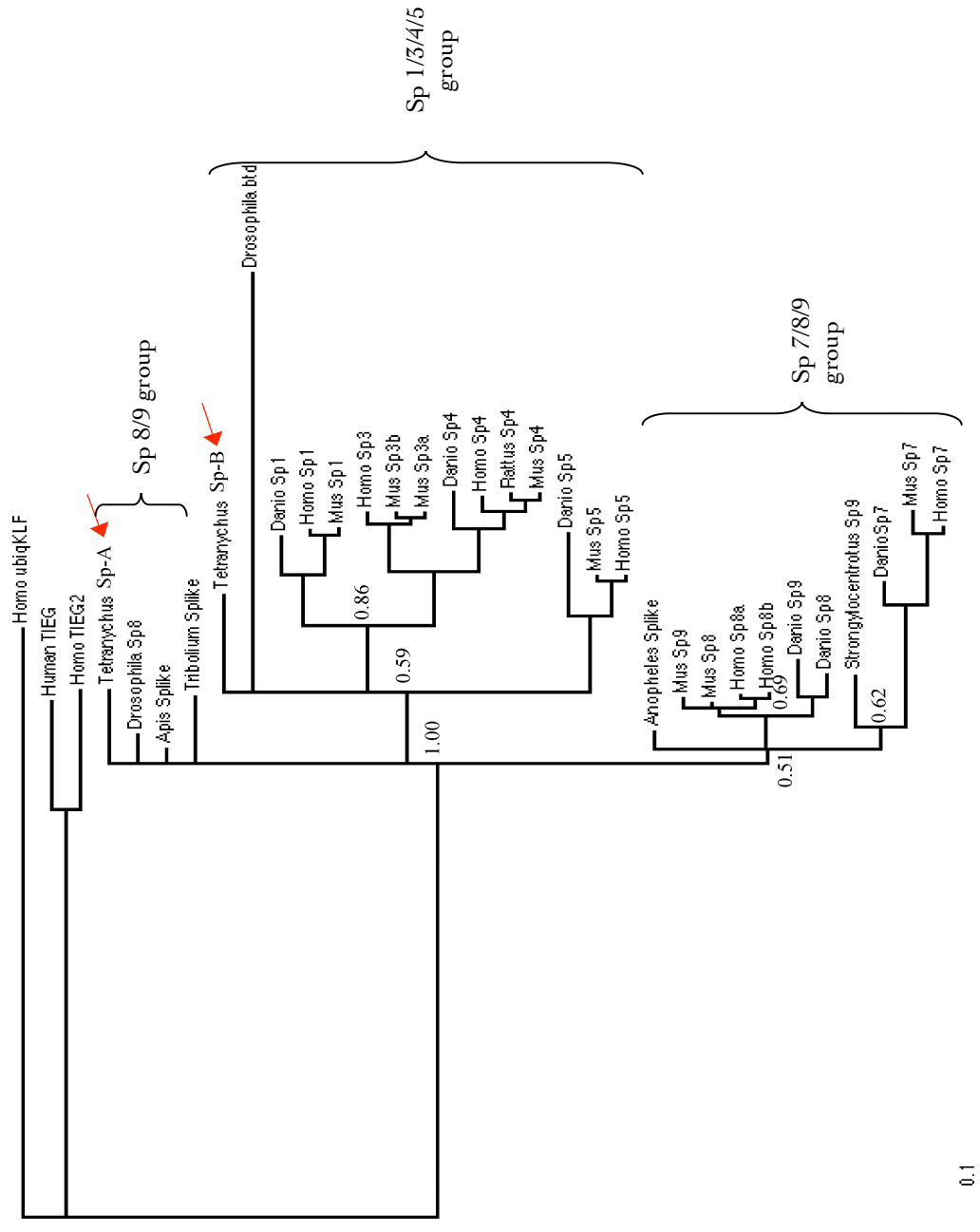


Figure 3.4.2a Bayesian consensus tree for analysis of putative *Tetranychus* *Sp* gene sequences. Orthology of the spider mite genes (red arrows) is not fully resolved, but *Tu-Sp-like1* appears to share closest affinity to *Sp8/9*-type genes, and *Tu-Sp-like2* is supported as an *Sp1/5/4*-type ortholog. *Drosophila btd* forms a long branch, consistent with its being a rapidly evolving gene formed by tandem duplication of an *Sp8*-type gene, hence prone to phylogenetic mis-placement (c.f. Chapter text regarding *Sp* gene family evolution) [Wimmer et al., 1995; Beermann et al., 2003]. *Homo ubiqKLF* and *TIEG* sequences are relatively long branches, forming an outgroup to the *Sp* family clade. Scale bar (0.1): expected base changes per site.

and is apparently more closely related to arthropod *Sp8* genes than to deuterostome *Sp7/8/9* sequences (Figure 3.4.2b). *Tu Sp-B* receives improved statistical support for belonging to a *Sp1/3/4/5* clade, and almost full support (0.99 probability) for a closer affinity to *Sp1/3/4* than the more divergent *Sp5* genes (Kawakami et al. 2004; Thorpe et al. 2005; Weidinger et al. 2005).

iii) *Tu Sp* gene orthologies and evolutionary context

Considering the phylogenetic evidence for gene orthology, *Tu Sp-A* will be referred to as '*Tu Sp8/9*', and *Tu Sp-B* as '*Tu Sp1/3/4*' from this point onwards. A summary of current ideas on *Sp* family evolution (Figure 3.4.2c) indicates an ancient tandem duplication giving two ancestral Urbilaterian *Sp* genes, one *Sp8/9*-type and one *Sp1/3/4*-type, feasibly - if further sequencing bears out their identity - orthologous to *Tu Sp-A* and *Tu Sp-B* (Kawakami et al. 2004) Wimmer, (pers.com.). Furthermore, the fast-evolving Dipteran gene *buttonhead* is proposed to be an *Sp8/9* paralog, arisen by tandem duplication (Estella, 2003; Schöck, 1999; Wimmer, 1995; Wimmer, 1996). Its higher relative rate of molecular evolution may explain both the long branch length of *Drosophila buttonhead* in Figure 3.4.2a and 3.4.2b topologies, and also its apparently spurious phylogenetic placement with *Sp1/3/4* rather than *Sp8/9* genes: it may have acquired, by chance substitution since duplication, a greater sequence similarity to *Sp1/3/4* genes.

3.4.3 Note on *Tetranychus Sp* gene expression

In spite of the minimal size of gene fragments so far cloned, I synthesised short DIG-AP labelled ssRNA probes from purified plasmids carrying ~170bp *Tetranychus Sp* gene inserts, to test for any ability, however negligible, to detect *Tu-Sp* mRNA expression *in situ*. As may be expected, using *Tu-Sp8/9* and *Tu-Sp1/3/4* RNA probes both arguably several hundred bases shorter than would be recommended for effective probe-target recognition, I only obtained what seem to be negative results for *in situ* hybridisation in whole mount spider mite embryos. With a *Tu-Sp8/9* probe I detected tentative transcripts on one side of some blastoderm stage embryos (7-9hr AEL), but no other *Sp8/9* expression was detectable at any other stage, calling the validity of the former into question as some kind of fixation or processing artefact. Details pertaining to *Tu-Sp* RNA probe synthesis, *in situ* hybridisation results and discussion of possible roles of *Sp* genes in spider mite development - considering remarkable functional conservation in other taxa - are to be found in **Appendix 6**.

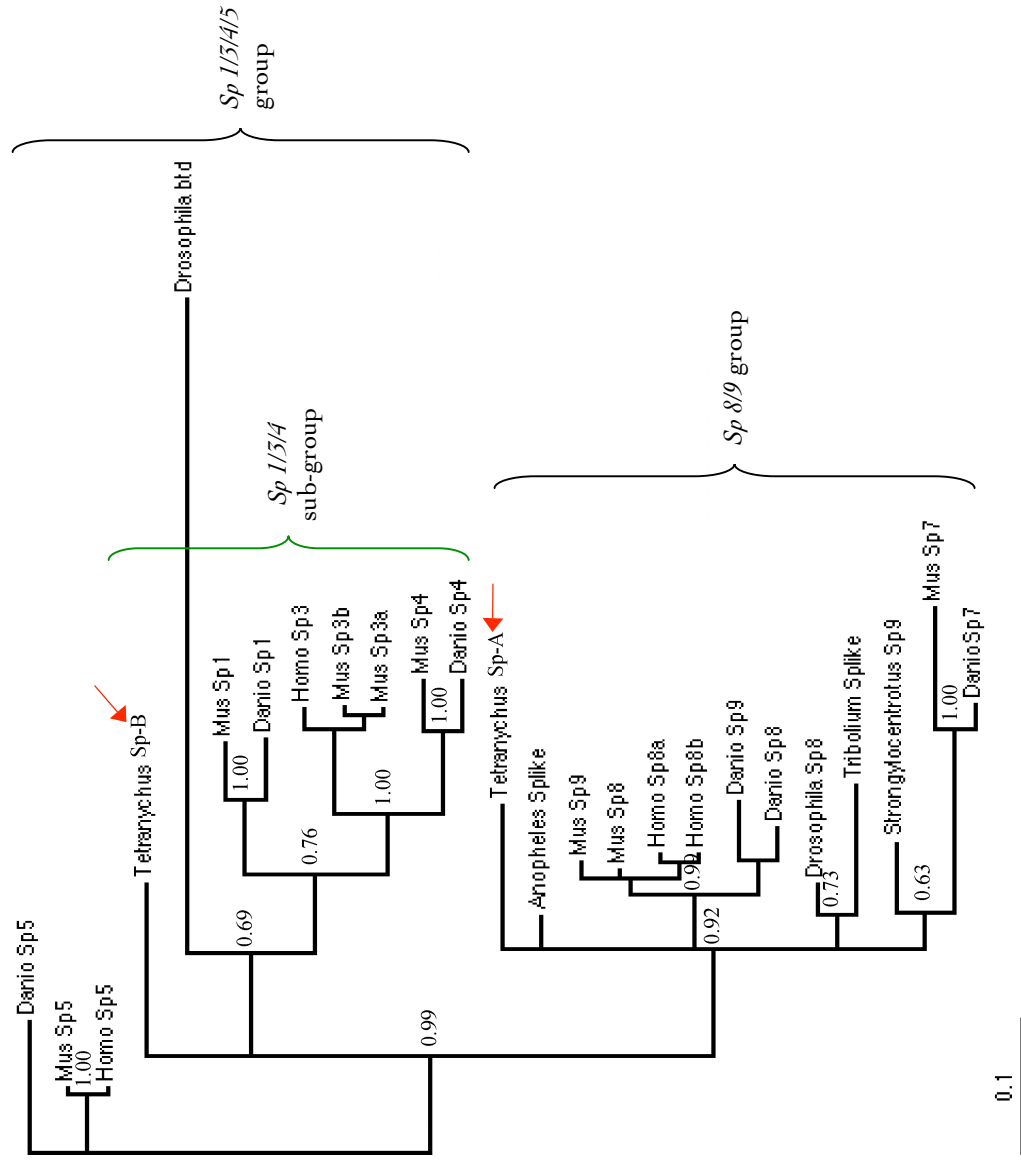


Figure 3.4.2b Bayesian consensus tree from analysis of putative *Tetranychus* *Sp* gene sequences (red arrows). The topology is un-rooted, without the outgroup from Figure 3.4.2a. Almost maximal (0.99) posterior probability supports the separation into *Sp 1/3/4* and *Sp8/9* clades. *Tu Sp-A* receives high (0.91) support as an *Sp8/9* gene, as do previously unassigned *Anopheles* 'Sp-like' and *Tribolium* 'Sp-like' genes. *Tu Sp-B* is of closer affinity to *Sp1/3/4* than *Sp5* genes, in keeping with the hypothesis that *Sp* genes evolved from an ancestral *Sp8*-type + *Sp1/3/4*-type gene pair, followed by divergence of *Sp5* from an original *Sp7*-type gene. The *Sp 7* genes are proposed to have evolved, along with *Sp6* genes, from *Sp 8* and *9* following a quadruple duplication event in the vertebrates [Kawakami et al., 2004; Wimmer pers. comm.]. Scale bar (0.1): expected base changes per site.

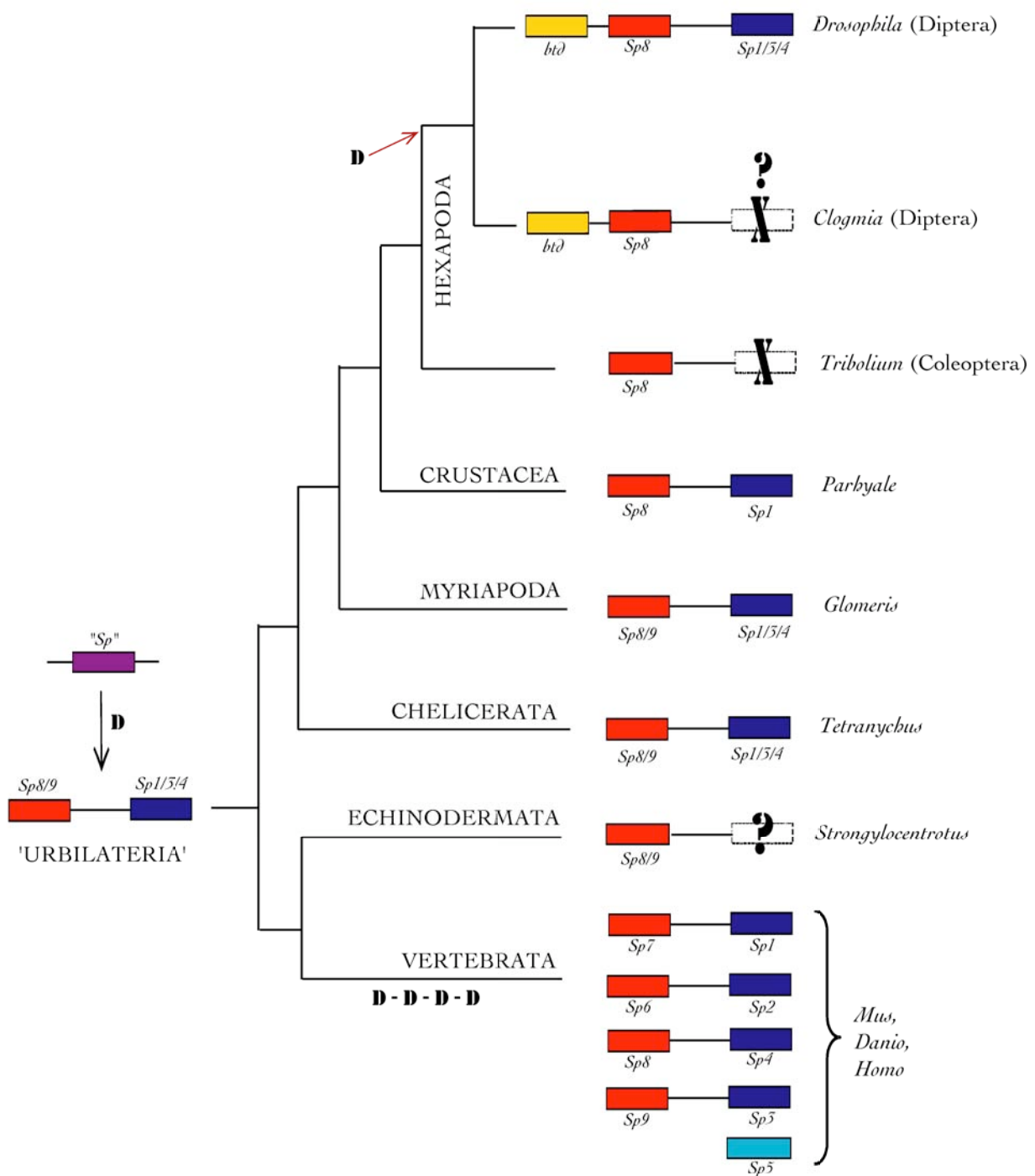


Figure 3.4.2c Model of *Sp* gene family evolution, based on available protostome (arthropod) and deuterostome data. Tandem duplication (**D**) of an original "*Sp-like*" gene produces one *Sp8/9*-type and one *Sp1/5/4*-type gene in an urbilaterian ancestor. The gene pair is retained in arthropod lineages, Diptera containing instances of apparent gene loss (**X**) and further tandem duplication of *Sp8* to give the gene *buttonhead* (*btδ*). It is not clear (?) whether the *Sp1/5/4* gene lost in *Tribolium* has also been lost independently in *Strongylocentrotus*. Four whole genome duplications occurred in the vertebrate lineage, giving rise to the two large sub-groups illustrated simply in the tree diagram.

Figure based on: Wimmer et al. (1995, 1996, pers. com.); Estella et al. (2003); Beermann et al. (2003); Janssen (pers. com.); Kawakami et al. (2004); Panopoulou & Poutska (2005); *Tribolium* genome project; my *Tetranychus*.

Conclusions to Chapter III

i) *Tetranychus Distal-less*

The results presented in this chapter have shown that the *Tetranychus urticae Dll* ortholog is transcribed during limb specification dynamically during proximo-distal limb patterning, and additionally in domains related to anterior CNS and PNS sensory organ development. *Tu-Dll* protein was not detected with a polyclonal *Precis Dll* antibody due to either spider mite epitope divergence or antibody degradation; however, I assume that mRNA transcript distribution reliably reflects localisation of the translated Dll transcription factor.

Tu-Dll in limb specification and proximo-distal patterning

First *Tu-Dll* expression at 21hrs AEL is associated with prosomal limb specification: Pp-L3 limb primordia are specified ventro-laterally just ahead of the chelicerae, proving the existence, at a molecular level, of delayed cheliceral segment formation that is synapomorphic for Chelicerata but was not completely clear from *Tetranychus* morphology alone (c.f. Chapter II). *Tu-Dll* expression confirms late specification of the L4 limb primordium and its developmental arrest until after hatching. The pattern of prosomal segmentation and limb formation - marked by *Tu-Dll* activity - broadly concurs with observations made in other acarids, in contrast with different recorded modes of opisthosomal segment delineation; this supports distinct segmentation mechanisms in anterior and posterior tagma. After affecting primordial limb specification, *Tu-Dll* expression dynamically affects proximo-distal limb development, defining the whole telopod and later, a subset of telopod segments marked by medial ring and distal tarsus mRNA transcription. A 'ring and tarsus' *Dll* pattern is most reminiscent of that seen in higher insects, although multiple rings are observed in outgrowing legs of spiders. Presumably species-specific *Dll* regulation occurs along proximo-distal limb axes, related to morphogenesis and/or sensory organogenesis unique to each lineage. Notably, *Tu-Dll* mRNA distribution consistently supports an appendicular affinity for the labrum: the labrum forms at the anterior midline from bilateral cell clusters in the ocular segment, and comes to have a distinct bilobed form with *Tu-Dll* expressed in restricted domains that correspond to 'distal' portions of two directly opposed, vestigial appendages.

Tu-Dll in neurogenesis

In addition to limb specification and proximo-distal patterning, *Tu-Dll* appears to function in early anterior CNS development, being expressed in a putative proto-cerebral domain that resolves to bilateral ocular segment domains related to innervation of the emerging labrum and/or later visual system patterning. *Tu-Dll* expression in both the CNS (e.g. proto-cerebrum, visual system) and PNS

(e.g. ‘empodium’ specification in growing limb buds), suggests a degree of functional conservation since proto-deuterostome divergence, as *Dll* is deployed in neurogenesis in a wide range of metazoans.

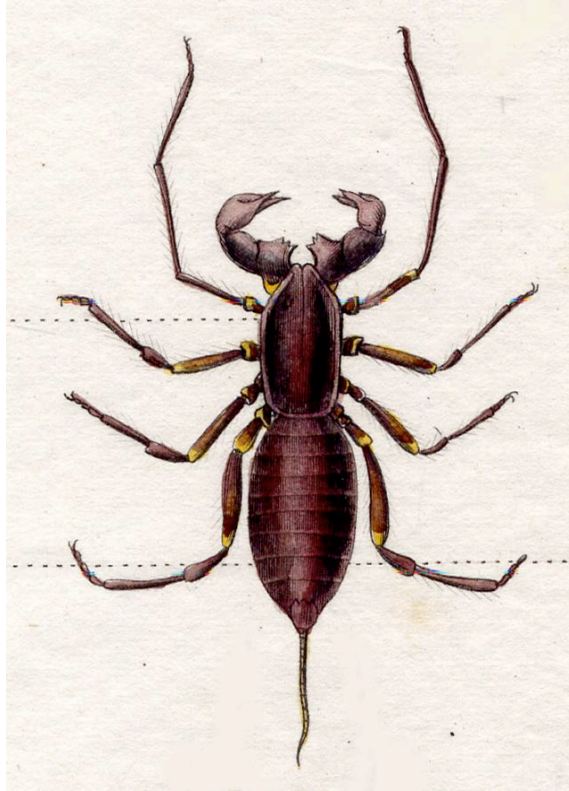
ii) *Tetranychus* *Sp* genes

As shown in sections 3.4 - 3.5 of this chapter, I obtained preliminary *Sp* gene fragments for spider mite orthologs of both an *Sp8/9* and an *Sp1/3/4* gene. These two *Tu-Sp* genes are predicted to correspond to the inferred ancestral Urbilaterian *Sp* gene pair, but as only 180bp of the conserved Btd and Zn finger domains is available for *Tu-Sp8/9* and *Tu-Sp1/3/4*, further sequence data are required to confirm these orthologies. Similarly, longer *Tu-Sp* gene fragments are needed for synthesis of adequate ssRNA probes for *in situ* hybridisation, as I could not detect *Tu-Sp* transcripts effectively with short, 180bp DIG-labelled RNA probes.

Sp gene activity (*Sp8/9* particularly) appears to be conserved in directing limb outgrowth and anterior CNS development in both arthropods and vertebrates, and furthermore there is evidence for *Sp* function in arthropod segmentation, limb allocation, limb allometry and PNS development. Thus, further work on spider mite *Sp* gene expression could reveal novel or conserved dynamic roles in limb and nervous system formation, and allow better assessment of how appropriate it is to assume that deep homology explains commonalities in arthropod and chordate *Sp* gene deployment.

CHAPTER IV

APPENDAGE SPECIFICATION IN *TETRANYCHUS URTICAE*: *WINGLESS*, *WNT* GENES AND *ENGRAILED*



Théliphone à queue

Introduction to Chapter IV

"I haven't failed, I've found ten thousand ways that don't work." - T. A. Edison

Why study *Tetranychus* *En* and *Wg/Wnt1*?

The segment polarity gene *engrailed* (*en*) is expressed in 14 single-cell wide stripes along the antero-posterior axis of the *Drosophila* blastoderm, activating the segment polarity gene *wingless* (*wg*) in cells anterior to the *en* row (c.f. Ch I, section 1.2.1). Auto-regulatory feedback loops between Hedgehog (Hh), induced by En signalling, and Wg signalling confers a sharp anterior limit to *engrailed* stripes, creating strictly defined organising centres at A-P parasegment boundaries. In *Drosophila*, both En and Wg signals are required for parasegmental organisation, but it is specific transduction of Wg signals that drives initial *Dll* expression at the 'early' *Dll-504* enhancer, providing positive regulatory input for limb primordium specification. (Temporal regulation of *en* and *wg* is clearly complex: once parasegments are established, *en* and *wg* are expressed independently of each other, and after limb specification, *Dll* transcription is driven from a downstream 'late' enhancer, independent of initial Wg cues.) During germband segmentation in arthropods other than *Drosophila*, *en* and *wg/Wnt1* orthologs are also consistently expressed in single-cell wide ectodermal stripes demarking anterior and posterior parasegment boundaries (c.f. Ch I, section 1.2.2). Conservation of En-Wg genetic circuitry in defining and maintaining parasegmental boundaries raises the possibility of a broadly conserved mechanism for *Dll* activation during primordial limb allocation along the antero-posterior axis, which occurs once germband segments are stable. I sought to obtain sequence and expression data for spider mite orthologs of *engrailed* and *wingless*, aiming to investigate their activity during early limb development in a chelicerate, distantly separated in phylogenetic terms from the model insect *Drosophila*, and hence broadening our appreciation of conservation vs. divergence in mechanisms of *Dll* regulation by segment polarity genes.

Tetranychus *Wnt* genes

In sections 4.1 of this chapter I report on extensive attempts to clone a *Tetranychus* *Wg/Wnt1* homolog, which unfortunately were not successful. A number of *Tetranychus* *Wnt* genes were recovered in an EST project at Agriculture Canada, and in section 4.2 I present sequence and phylogenetic analysis for three of these genes: their identity is not fully resolved, but other aspects of consensus tree topologies point to a more diverse complement of Wnt family members in the ancestral Ecdysozoan lineage than previously thought. Further work is clearly required to recover full length spider mite

Wnt genes, confirm the loss of a *Wnt1* ortholog, and to demonstrate gene expression patterns for those *Wnt* paralogs that are present. Antibodies against Wingless and the Wnt signal transduction factor Armadillo/ β -catenin were tested on whole-mount spider mite embryos at different stages of embryogenesis, but no protein was detected. Given the likelihood of conserved Wnt gene deployment in establishing parasegmental boundaries and/or mediating cell-cell interactions in other segmental and non-segmental contexts, failure of antibodies is not considered indicative of absent Wg/Wnt activity, but rather epitope divergence or reagent degradation.

***Tetranychus engrailed* genes**

In section 4.3 of this chapter I present sequence data for two *Tetranychus engrailed* paralogues cloned by degenerate PCR and extended by inverse PCR methods: the genes were provisionally termed *Tu-en1* and *Tu-en2*. Almost 1.2Kb was recovered for *Tu-en1*, significantly less (360bp) for *Tu-en2*, and their orthology to *engrailed* class genes was demonstrated by Bayesian inference. Engrailed duplication is found throughout the arthropods and in several other independent metazoan lineages, associated with varied divergence in sequence and expression that reflects para- or sub-functionalisation. To characterise *Tetranychus en1* and *en2* mRNA distribution separately, I synthesised several ssRNA probes between 350bp and 985bp in size for each paralog. However, transcript detection was unsuccessful, for possible reasons discussed in the text of section 4.4. A Canadian research group visualised mRNA expression of a *Tetranychus engrailed* gene in late stage embryos, revealing a typical pattern of *en* activity in posterior domains of segmental appendages: possible gene expression during segmentation, limb specification and other roles such as neurogenesis remain to be shown for this, and the second *Tu-en* paralog. In whole mount spider mite embryos, I tested monoclonal antibodies against two *Drosophila* En paralogs that are known to cross-react in other species. No protein was detectable, and comparing spider mite En1 and En2 homology domains with those of other taxa showed that epitope divergence may explain this lack of cross-reactivity.

4.1 Cloning *Tetranychus urticae* *Wnt* genes

4.1.1 Degenerate PCR screen for a *Tetranychus* *Wg/Wnt1* ortholog

i) Negative results: failed target amplification

I could not recover a spider mite *Wnt1* ortholog by degenerate PCR, using either genomic DNA, embryo cDNA or total RNA template. I tested combinations of 5 forward and 8 reverse primers, designed to target conserved domains within either all *Wnt* family genes (ADP-Wg; CCK-Wg F1,R1,R2) or *Wnt1/Wg* homologs specifically (CCK-Wg F1; CsWg): see Figure 4.1.1a for domain specificity details, and Appendix 2.3 for all primer sequences. I designed 4 (CCK) primers, 4 (CsWg) were based on those in Damen (2002), and 5 (ADP) were kindly given by A.D. Peel (Damen 2002; Sidow 1992). Although PCR reactions were carried out using multiple primer combinations, thermal conditions, polymerase forms, Mg^{2+} ion concentrations ($[Mg^{2+}]$) and DNA/RNA templates (as described below), no target DNA fragments were amplified. Methods and conditions tested (c.f. Ch VIII, section 8.2 for programs) include the following:

- Standard PCR on genomic DNA (embryonic and nymph/adult) and embryonic cDNA, with 2.5mM or 3.0mM $[Mg^{2+}]$.
- Touchdown PCR (8x 1°C initial decrease cycles), annealing at 50°C, 51°C, or 52°C and with 2.5mM or 3.0mM $[Mg^{2+}]$.
- Touchdown PCR with 10x decrease increments (60-50°C).
- RT-PCR on embryo or nymph/adult total RNA, standard or touchdown PCR (8x or 10x temperature decrease increments) after first strand synthesis reaction.
- Gradient (50-58°C) and touchdown PCR (50°C annealing, 2.5mM $[Mg^{2+}]$) on cDNA using Titanium Taq (polymerase more active than standard Taq).
- Re-PCR of low conc. PCR/RT-PCR products, or 'gel stab' material obtained by pipette tip stabbed into an agarose gel to precisely sample faint DNA bands.
- Nested PCR after touchdown PCR with gDNA, cDNA template or after RT-PCR product synthesis.

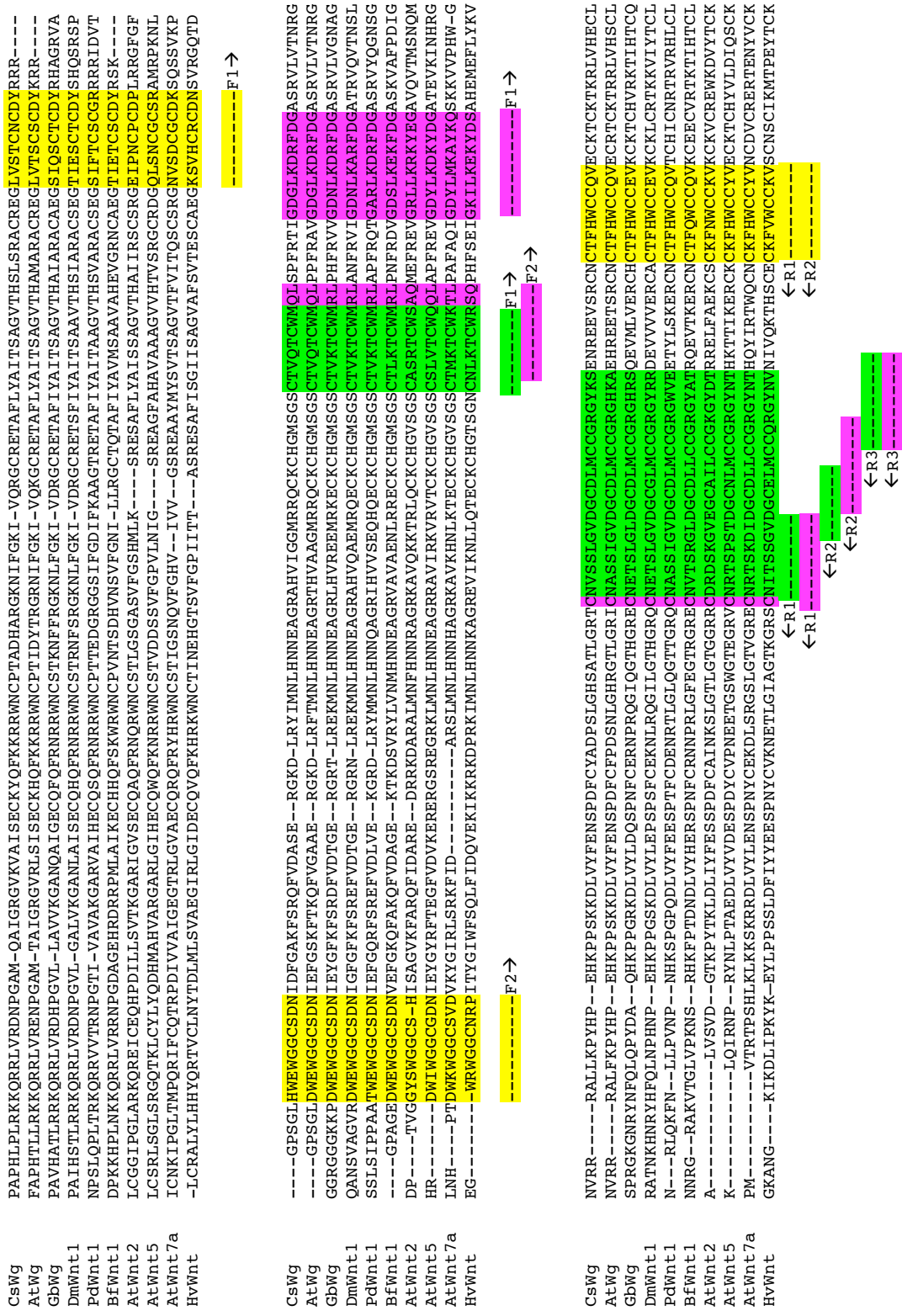


Figure 4.1.1a Multiple Wnt protein alignment (primarily Wnt1 orthologs), to show domains targeted by Wg/Wnt degenerate PCR primers. **Yellow:** CsWg primers, specifically targeting spider and arthropod Wg genes. **Green:** CCK-Wg primers, F1 designed against arthropod Wg, and F2, R1, R2 against conserved Wnt family sequences. **Purple:** ADP-Wg primers, targeted against conserved Wnt family domains.

ii) Negative results: spurious amplification

Predicted target DNA fragment sizes are 275-350bp with ADP/CCK primers and 455-500bp with CsWg primer pairs (Figure 4.1.1a). Multiple DNA fragments between 200bp and ~900bp in size were amplified under the following conditions:

- a) degenerate PCR with ADP primers (Figure 4.1.1b, i)
- b) RT-PCR and Ti Taq PCR on cDNA with CsWg primers (Figure 4.1.1b, ii & iii)
- c) nested PCR on cDNA with CCK/CsWg primers (Figure 4.1.1b, iv).

Amplification of multiple products with molecular weights either above or below the expected size was indicative of non-specific primer binding, which was indeed confirmed: Cloning and sequencing yielded only non-target or 'junk' sequence (vector material, bacterial DNA, primer artefacts), proving that mis-priming or non-specific binding within the genome had indeed caused amplification of random DNA sequences, unrelated to *Wnts*.

4.1.2 *Tetranychus* *Wnt* genes from an EST screen

i) *Tu-Wnt* sequences

A *Tetranychus urticae* EST database exists at Agriculture Canada (Ontario), and several putative *Wnt* gene fragments have been identified by BLAST sequence identity comparison. Three sequences assigned to the *Wnt* family were donated to me for further analysis; for simplicity their EST identity codes (#144, #145 and #156) were replaced by provisional gene names *Tu-WntA*, *Tu-WntB* and *Tu-WntC* respectively. Removal of poor sequence data left 303bp *Tu-WntA*, 423bp *Tu-WntB* and 630bp *Tu-WntC*, all three coding for relatively cysteine-rich polypeptides (Tu-WntA, 5% cys; Tu-WntB, 5.7%; Tu-WntC, 7.6%): for nucleotide and amino acid sequence data see Figure 4.1.2a. *Tu-WntA* and *Tu-WntB* code for proteins with 69.2% amino acid identity, including 4 conserved cysteines and 1 conserved asparagine residue (Figure 4.1.2a,iii). *Wnt* family proteins are typically cysteine-rich, enabling ligand binding *via* covalent disulphide bridges to the extracellular cysteine-rich domains (CRDs) of Wnt receptors such as the serpentine trans-membrane receptor Frizzled. Numerous possible sites of asparagine-linked glycosylation are also evident (Figure 4.1.2a), a modification that labels Wnt ligand proteins for secretion (Eisenberg et al. 1992; Jones and Jomary 2002; Nusse and Varmus 1992).

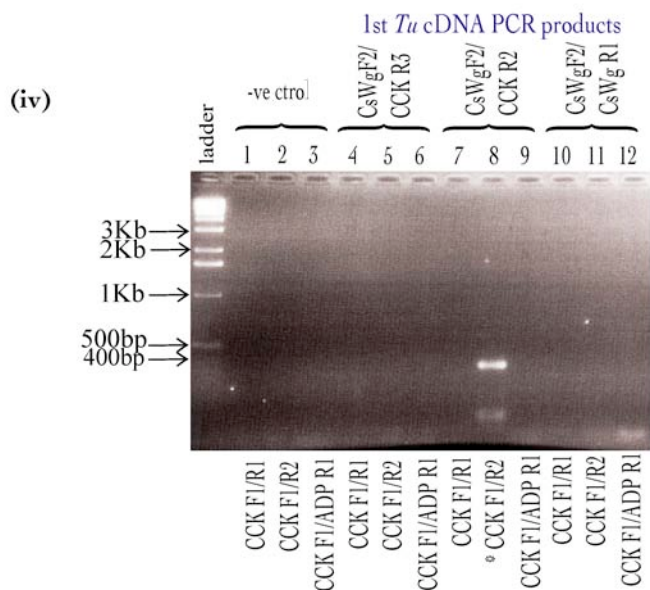
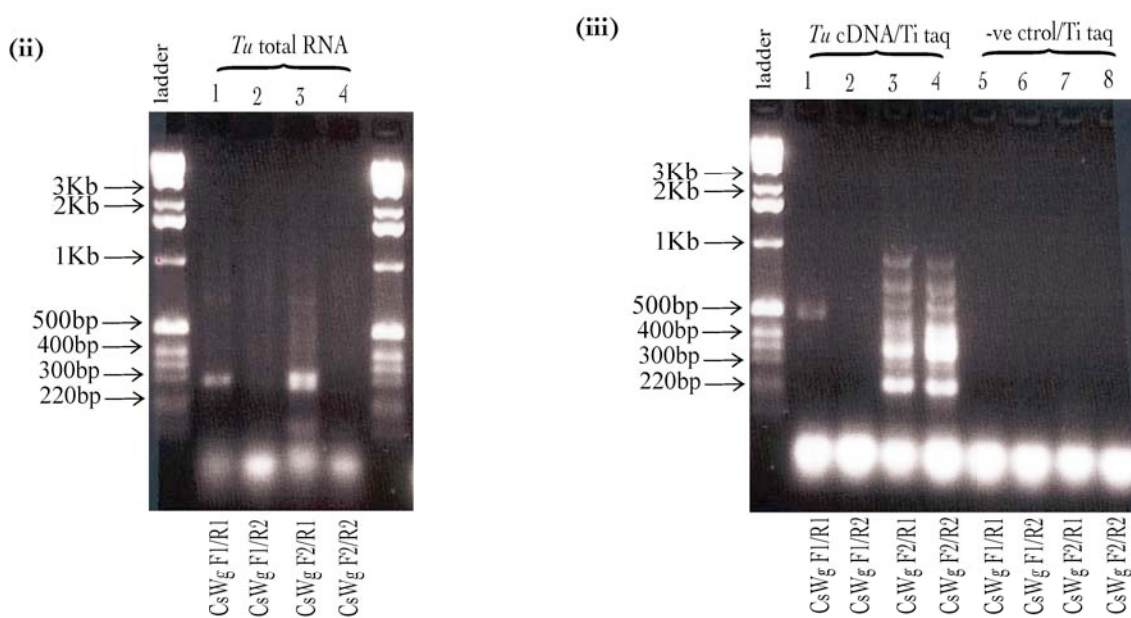
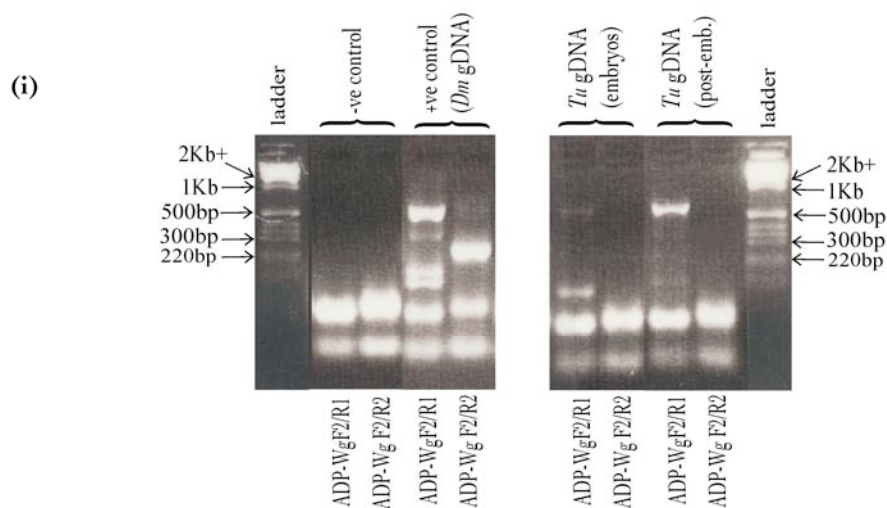


Figure 4.1.1b Amplification of spurious DNA fragments during degenerate PCR screen for *Tetranychus* *Wnt* genes.

- (i) Touchdown PCR on embryonic and nymph/adult genomic DNA template, with ADP primers.
- (ii) RT-PCR with CsWg primers, on *T. urticae* total RNA.
- (iii) Titanium taq PCR on *Tetranychus* cDNA, with CsWg primers (Wg/Wnt1 or Wnt family specific).
- (iv) Nested PCR after preliminary PCR on *Tetranychus* cDNA, with primers as labelled.

(i) *Tu-WntA*

```
1/1                               31/11
GAG CTC GCG CGC CTG CAG GTC GAC ACT AGT GGA TCC AAA GAA TTC GGC ACG AGG CGT AAA
glu leu ala arg leu gln val asp thr ser gly ser lys glu phe gly thr arg arg lys
61/21                               91/31
GAG TGT GAA TTT CAG TTC CGT AAT CGT CAA TGG AAT TGT CCA TCA ACG CGA AAG TCG ATG
glu cys glu phe gln phe arg asn arg gln trp asn cys pro ser thr arg lys ser met
121/41                               151/51
AGG AAG ATC CTT TTG AAA GAT ACT CGG GAG ACT GGT TTC GTT CAT GCT ATC ACA GCA GCT
arg lys ile leu leu lys asp thr arg glu thr gly phe val his ala ile thr ala ala
181/61                               211/71
GGA ATG ACA CAT AGC TTA GCT AAA GCG TGT AGC CAA GGA ACT CTT CTC GAT TGT TCT TGC
gly met thr his ser leu ala lys ala cys ser gln gly thr leu leu asp cys ser cys
241/81                               271/91
CTT ATG ATG TCC TCT CCT CAA TCA GAT AAT TCA TTG TCT CAA GAT TTT GAT TTT CCG CCG
leu met met ser ser pro gln ser asp asn ser leu ser gln asp phe asp phe pro pro
301/101
GTT
val
```

(ii) *Tu-WntB*

```
1/1                               31/11
CAG TTA TTC GTA AAA ATT CTT TGT CTA ATC TCA ATC ATC GAA TGT CGC ACG AGT GAA TCA
gln leu phe val lys ile leu cys leu ile ser ile ile glu cys arg thr ser glu ser
61/21                               91/31
TTT TGG GGT ACG GGA ATG AAA CTT GTT TTA GAT CCG AAT AAA ATC TGC CGT AAA GCA GGT
phe trp gly thr gly met lys leu val leu asp pro asn lys ile cys arg lys ala gly
121/41                               151/51
CGA TTG AAA AGT AAA CAA ACA AGT ATT TGT GAA AAA GGA CCG GAG ATC GTT CGG GAA ATA
arg leu lys ser lys gln thr ser ile cys glu lys gly pro glu ile val arg glu ile
181/61                               211/71
ACC AAA GGT GCT AAA ATT GCC CGT AAA GAG TGT GAA TTT CAG TTC CGT AAT CGT CAA TGG
thr lys gly ala lys ile ala arg lys glu cys glu phe gln phe arg asn arg gln trp
241/81                               271/91
AAT TGT CCA TCA ACG CGA AAG TCG ATG AGG AAG ATC CTT TTG AAA GAT ACT CGG GAG ACT
asn cys pro ser thr arg lys ser met arg lys ile leu leu lys asp thr arg glu thr
301/101                               331/111
GGT TTC GTT CAC GCT ATC ACA GCA GCT GGA ATG ACA CAT AGC TTG GCT AAA GCG TGT AGC
gly phe val his ala ile thr ala ala gly met thr his ser leu ala lys ala cys ser
361/121                               391/131
CAA GGA ACT CTT CTC GAT TGG TCT TGC CTT ATG ATG CCT CTC CTC AAT CAG ATA ATT CAT
gln gly thr leu leu asp trp ser cys leu met met pro leu leu asn gln ile ile his
```

(iii)

```
TuWntA 5'-----ELARLQVDTSGSKEFGTRRK
TuWntB 5'QLFVKILCLISIIECRTSESFWGTGMKLVLDPNKICRKAGRLKSKQTSICEKGPEIVREITKGAKIARK
                                         *      **

TuWntA ECEFQFRNRQWNC PSTRKSMRKILLKDTRETGFVHAITAAGMTHSLAKACSQGTLLDCSCLMMSSPQSDN
TuWntB ECEFQFRNRQWNC PSTRKSMRKILLKDTRETGFVHAITAAGMTHSLAKACSQGTLLDWSCLMPLLNQIIF
*****

TuWntA SLSQDFDFPPV 3'
TuWntB H----- 3'
```

Figure 4.1.2a
Page 1 of 2

(iv) *Tu-WntC*

```
1/1
TGC AGG TCG ACA CTA GTG GAT CCA AAG AAT TCG GCA CGA GGC TCA ACA AAC CTT ACT CAA
cys arg ser thr leu val asp pro lys asn ser ala arg gly ser thr asn leu thr gln
61/21
TGC TCT TCA ACC TTC ACC GAC CCC GAA ATT AGA GCC AAA AGT CTT CAC ATA AAA AGA GAT
cys ser ser thr phe thr asp pro glu ile arg ala lys ser leu his ile lys arg asp
121/41
GCT CGT AAA ATT AAA TTA TCA AAT CGT CAA AAT GGT TCC GGA TAC AAG CCA CCA AGT GAT
ala arg lys ile lys leu ser asn arg gln asn gly ser gly tyr lys pro pro ser asp
181/61
TCG GAT CTT GTT TAC CTT CAC TCA TCG CCA GAC TAT TGT GAA AGG GAT GAA AAA TCA GGT
ser asp leu val tyr leu his ser ser pro asp tyr cys glu arg asp glu lys ser gly
241/81
TCC CCT GGA ACC CAT GAT AGA TTT TGT AAT GGA ACC TCT AAG GGT GCT GAT GGT TGT GAT
ser pro gly thr his asp arg phe cys asn gly thr ser lys gly ala asp gly cys asp
301/101
ATA CTT TGT TGT AAT AGA GGT TTT ACT CGA CGA TTT GAG ACT GTT AAG GAA AAA TGT AAT
ile leu cys cys asn arg gly phe thr arg arg phe glu thr val lys glu lys cys asn
361/121
TGT AAA TTC TTT TGG TGT TGT CAA GTT GAA TGT GAC GAA TGT ACT CAT CAG ATT GAG ATC
cys lys phe phe trp cys cys gln val glu cys asp glu cys thr his gln ile glu ile
421/141
AAC ACT TGC TTG TGA ATT GTT GAG TCA ACA AAT TCT AGC AAC GAC AAG AGT TCT TAA AAG
asn thr cys leu -X-
481
AAA ACA AAG TTC AAG CAA AAA AGC CCA AGT TAA AGC CAA ACC AGG CCT AAC ACA GCA AGA
541
AAT CAA AGG GGA AAT AAA GTT TTT AGA GGA AAA GGC AAA ATG GAA TCA AAC AAT CAT CAA
601
TTT CAT CAA CTT GAA TTA CTC TTT TTC TTT
```

(v)

```
Tu-WntC 5' CRSTLVDPKNSARGSTNLTQCSSTFTDPEIRAKSLHIKRDARKIKLSNRQNGSGYKPPSDSDLVYLHSS
PDYGERDEKSGSPGTHDRFCNGTSKGADGCDILCCNRGFTRRFETVKEKCNCKFFWCCQVECDCTHQI
EINTCLX-3'
```

Figure 4.1.2a Nucleotide and amino acid sequences for EST-derived, putative *Tetranynchus Wnt* genes. All three *Tu-Wnts* are relatively cysteine-rich (*Tu-WntA*, 5%; *Tu-WntB*, 5.7%; *Tu-WntC*, 7.6%), potentially enabling Wnt ligand binding *via* disulphide bridges to extracellular cysteine rich domains (CRDs) of Wnt receptors. The presence of asparagines residues may indicate sites for glycosylation, a covalent modification allowing Wnt secretion as a glycopeptide. [Jones & Jomary, 2002; Eisenberg et al, 1992; Nusse & Varmus, 1992]

(i) *Tu-WntA*, 303bp CDS (101aa) from EST no. 144. (ii) *Tu-WntB*, 423bp CDS (141aa) from EST no. 145. (iii) Comparative alignment of *Tu-WntA* and *Tu-WntB* protein sequences: 69.2% shared identity (63 of 91aa present in both), including 4 conserved cysteines (yellow) and 1 asparagine (blue), possibly functional in Wnt binding and secretion respectively (grey: un-conserved C or N residues). (iv) *Tu-WntC*, 630bp (435bp 3' CDS + 194bp UTR) from EST no. 156. (v) *Tu-WntC* polypeptide sequence. Blue: possible site of asparagine-linked glycosylation. Grey: cysteine residue, associated with ligand binding.

ii) Multiple Wnt gene alignment

Alignment of *Tu*-Wnts with Wnt paralogs from multiple metazoan taxa reveals a CRD comprising 22 conserved cysteine residues (Figure 4.1.2b,i-iv). *Tu*-WntA and *Tu*-WntB amino acids that were alignable with homologous residues in the other Wnt sequences align with the Wnt CRD N-terminal region, from cys 1-5 (Figure 4.1.2b,ii); *Tu*-WntC protein sequence aligns with the C-terminus, cys 7-22 (Figure 4.1.2b,iii). Of the numerous asparagine (N) residues coded for by *Tu*-Wnt sequences, only one in each of *Tu*-WntA, *Tu*-WntB and *Tu*-WntC is shown to be conserved relative to other metazoan Wnts (Figure 4.1.2b,ii&iii). Consistently conserved asparagine sites suggest a function in glycosylation, significant in mediating Wnt secretion (Logan 2003; Nusse and Varmus 1992; Wodarz and Nusse 1998). Across all Wnts in the *Tu*-WntA/B alignment, 12 of 73 (16%) amino acids are 100% conserved, and in the *Tu*-WntC alignment 21 of 86 (24%) are consistently conserved (Figure 4.1.2b,ii-iii).

4.1.3 Phylogenetic analysis of *Tu*-Wnt genes

i) *Tu*-WntA and *Tu*-WntB affinity

Incomplete resolution by Bayesian inference

By Bayesian inference I analysed homologous coding sequence from *Tu*-WntA, *Tu*-WntB and 22 protostome and deuterostome *Wnt* orthologs (see Chapter VIII for method). In general the resultant consensus tree (Figure 4.1.2c) is poorly resolved, presumably due to limited phylogenetic signal within the short (219bp) section of coding sequence compared. Nevertheless, a *Wnt5* clade receives almost maximal posterior marginal probability (0.99), supporting the hypothesis that it is an ancestral *Wnt* sub-family, present before Cnidarian-Bilaterian¹⁵ divergence in the Pre-Cambrian (Hobmayer et al. 2000; Kusserov et al. 2005; Llimargas and Lawrence 2001; Prud'homme et al. 2002; Schubert et al. 2001). *Tu*WntA and B genes tentatively group with *Wnt1* orthologs, but the relationship receives low statistical support (indicating easy collapse into alternative topologies), and the inclusion of *Homo*Wnt9a and *Achaea*Wnt2 genes within the proposed '*Wnt1*' clade also indicate low robustness (Figure 4.1.2c). Extension of *Tu*-WntA and *Tu*-WntB sequences is clearly required for analysis of a larger character set and confirmation of these genes' affinities.

¹⁵ The term 'Bilateria' is used with the caveat that certain Cnidaria (e.g. anthozoan *Nematostella vectensis*) display bilateral symmetry, including axial *Wnt*, *Hox* and *TGF-β* transcription patterns and elements of morphology (Finnerty et al. 2004; Hobmayer et al. 2000; Kusserov et al. 2005).

Cswg	VTNDLVYFNSP	DFCYADPSL	GHSA	TLGR	TGNVSS	LVGDG	DL	MC	CGRGYK	SENRE	EVSR	NC	TH	WC	CO	VE	CT	CK	TK	RL	VH	EC
Gmwg	VSNDLVYFNSP	DFCERD	TKLGP	PTGR	CFNC	DTSL	VG	DL	---	---	---	---	---	---	---	---	---	---	---	---	---	
TLwnt1	ISNDLVYFNSP	NPNC	EKN	TKLGP	TQGR	TCN	DTSL	VG	DL	MC	CGRGYK	TE	KE	VE	VR	CA	CT	---	---	---	---	
TDwg	MNDLVYFNSP	SPN	CDRNT	KLGT	QGR	CTN	DTSL	VG	DL	MC	CGRGYK	TE	KE	VO	ER	CA	CT	---	---	---	---	
Gbwg	VGNDLVYLO	SPN	PCERN	PROQ	TQGR	HC	ECN	DTSL	GL	DC	MC	CGRHS	Q	EV	ML	VE	CH	TF	HW	CE	VK	
Dmwnt1	VNDLVYLB	SP	SPCE	KN	RLG	ITG	TH	GR	QCN	DTSL	VG	DL	MC	CGRGYK	RD	VE	VV	VE	RA	CT	FW	
PdWnt1	QNDLVYF	FE	SP	TC	EN	RL	TG	ITG	TH	GR	QCN	AS	I	GV	DC	DL	MC	CGRGW	VE	ET	YL	
CeWnt1	VVKDLVIM	TP	SP	DC	ES	PL	GI	TG	KR	Q	CT	LAP	AD	C	S	L	V	---	---	---	---	
SpWnt1	MRNDLVYFNSP	DFCERN	DKF	GP	PTGR	EC	N	AT	SL	VG	DL	MC	SRG	S	T	TE	I	K	V	K	R	
BfWnt1	FPDDLVI	HERSP	PNR	PNR	PL	GF	TGR	EC	N	AT	SL	VG	DL	MC	SRG	S	T	TE	I	K	V	
HSnt1	VNDLVYF	HERSP	NPNC	T	SG	RI	TAG	T	AG	C	AN	SS	P	AL	D	G	CE	LL	CC	GRG	H	
TuWntC	LSNDLVYLS	SPD	CER	DEK	SG	PT	DR	FC	N	CT	SG	K	A	D	G	CE	DL	MC	NRG	F	R	
AtWnt2	MSNDLI	YFNSP	DP	FC	AI	N	KS	L	G	T	G	RR	C	R	D	S	K	G	V	E	G	
PvWnt2	INQDLVYFT	SP	DY	C	I	K	I	E	I	S	G	L	T	A	G	R	C	K	S	G	L	
CeWnt2	INDDLVFM	D	SP	DY	C	R	H	S	G	T	L	T	A	G	R	C	K	S	G	A	E	
McWnt3	INRDLVYMA	S	P	DY	C	K	R	N	T	V	S	L	T	S	R	I	C	N	T	S	P	
HSwnt3	VEKDLVYFNSP	NPNC	EP	ET	PG	TGR	EC	N	AT	SL	VG	DL	MC	CGR	N	T	H	T	E	R	K	
BfWnt4	SKDLVYLV	SP	DF	CE	ED	L	K	V	S	G	L	T	H	C	R	ON	K	T	S	A	I	
PdWnt4	QKDLVYLV	AS	P	DF	C	ER	D	P	K	T	G	A	L	T	H	C	R	ON	K	T	S	
AtWnt5	INHDLVYD	ES	P	DP	C	V	R	T	P	K	S	G	M	T	G	R	C	N	T	S	P	
AmWnt5	INRDLVYLD	SP	N	C	L	P	N	E	T	L	S	G	L	T	H	C	R	I	C	N	T	
Dmwnt5	INKDLI	L	D	ES	P	D	W	C	R	N	S	A	L	H	P	T	G	R	C	N	T	
BfWnt5	LI	KDLVYLN	K	SP	DY	C	N	A	D	P	T	I	S	G	L	T	H	C	R	E	C	
HeWnt5	VNDLVYPO	S	DP	CY	P	D	V	T	G	S	L	T	H	C	R	E	C	N	T	S	P	
SpWnt6	VNDLVYMS	H	ES	H	D	EP	R	K	S	G	L	T	H	C	R	E	C	N	T	S	P	
Cswnt7a	PQWDLVY	EN	SP	NY	C	ES	D	T	V	G	S	L	T	H	C	R	E	C	N	T	S	
Gmwnt7	PSAELVH	I	H	K	SP	N	C	I	E	D	K	S	G	L	T	H	C	R	E	C	N	
AmWnt7a	AI	TEL	V	F	L	P	SP	NY	C	E	D	L	V	O	G	S	L	T	H	C	R	
SpWnt7	PVKHLV	H	R	SP	NY	C	E	D	L	V	O	G	S	L	T	H	C	R	E	C	N	
BfWnt7a	PVRDLVY	I	E	K	SP	N	C	E	D	L	V	O	G	S	L	T	H	C	R	E	C	
BfWnt8	MYRDLVFL	E	D	SP	C	R	E	N	L	T	V	G	R	E	L	G	R	K	Y	E	K	
HSwnt8a	MDKELI	F	LE	S	P	DY	C	T	C	N	S	S	L	T	H	C	R	E	C	N	T	
HSwnt9a	STTELVLH	D	SP	S	PL	C	A	--	GR	F	S	P	T	AG	R	CH	E	---	KN	CS	I	
EptWnt9	SMTDLVY	DE	S	P	S	Y	C	R	A	--	S	R	S	P	T	AN	R	S	CO	G	---	
TLwnt10	QNDLVY	Y	ERSP	CF	ET	CD	P	V	F	G	S	R	H	C	N	R	SS	Q	D	I	G	
TDwnt10	QSNDLI	Y	ERSP	NPNC	E	K	A	S	V	GP	T	G	R	C	N	T	S	G	L	T	H	
Dmwnt10	QNSLFI	Y	Q	R	SP	N	PC	R	D	L	G	A	D	I	O	G	T	V	G	R	K	
BfWnt10	VNDLVYF	OR	SP	DC	R	N	E	L	E	T	P	G	T	R	I	C	N	T	S	T	G	
BfWnt11	KRKLFI	V	EN	SP	N	C	M	V	N	R	K	S	G	T	G	T	C	N	T	S	P	
HomeWnt11	HRPELVI	Y	FN	SP	DC	M	K	N	E	K	V	G	H	G	T	O	D	R	C	N	T	
HomeWnt13	ATQDLVI	Y	FN	SP	DY	C	V	L	D	K	A	A	S	L	T	G	R	C	V	S	K	
HydraWnt	FLYDPI	Y	E	E	S	P	N	C	V	K	N	E	T	L	G	I	A	T	G	R	S	

Figure 4.1.2b Multiple protein sequence alignment for Wnt genes, *Tetranychus* Wnts in bold. (i) Alignment of Tu-Wnts with *Cupicinnus* Wnt1 CRD (209aa) to show relative location of CDS used in subsequent Bayesian analyses. (ii) Wnt alignment (26 sequences, 73 amino acids) including Tu-WntA and Tu-WntB; domain reflects homologous CDS used in Bayesian analysis. (iii) Wnt alignment (43 sequences, 127 amino acids) including Tu-WntC; reflects CDS used in Bayesian analysis. Asterisk: 100% conserved residue (or in all but one sequence). Grey: conserved residue. Yellow: conserved cysteine, for ligand binding. Blue: conserved asparagine, for mediating glycosylation and ligand secretion. Species abbreviations: Cs - *Cupicinnus* *saltii*; Gm - *Glomeris* *marquandii*; Tl - *Triplo* *longicaudatus*; Td - *Thermobia* *domestica*; Gb - *Grigulus* *bimaculatus*; Hs - *Homo* *sapiens*; Sp - *Strongylocentrotus* *purpuratus*; Bf - *Branchiostoma* *flavescens*; Pv - *Patella* *vulgaris*; Pd - *Platynereis* *dumerilii*; Ce - *Caenorhabditis* *elegans*; Tu - *Tetranychus* *urticae*; Am - *Apis* *melifera*; At - *Arabidopsis* *thaliana*; Mc - *Mytilus* *edulis*; He - *Helicoverpa* *armigera*; Ept - *Eptatetranychus* *sp.*

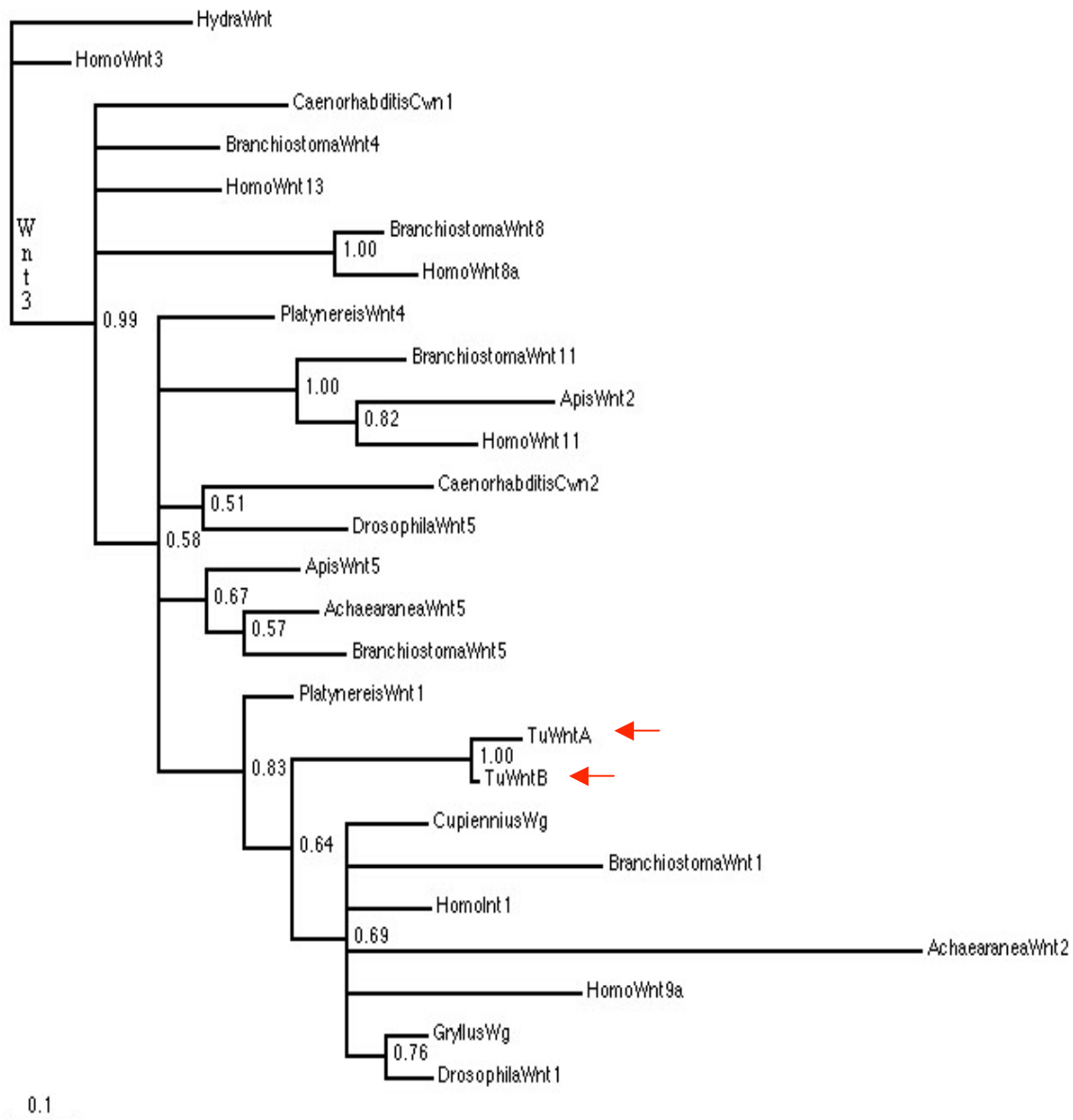


Figure 4.1.2c Bayesian inference consensus for analysis of *Tetranychus urticae* *WntA* and *WntB* aligned CDS (219bp) with 22 Wnt orthologs from other taxa. Resolution is incomplete due to limited phylogenetic signal within the short CDS available, although almost maximal p.p. support is retrieved for *Wnt5* genes as ancestral within the *Wnt* superfamily. *Tetranychus Wnts*: red arrows.

Wnt gene loss in the Ecdysozoa

Loss of *Wnt2*, *Wnt3*, *Wnt4*, *Wnt8* and *Wnt11* sub-family genes from the Ecdysozoa has been proposed on the basis of genome data from *Drosophila melanogaster*, *Anopheles gambiae* and *Caenorhabditis elegans* only (Kusserov et al. 2005; Prud'homme et al. 2002). However, the Figure 4.1.2c topology identifies *Apis* 'Wnt2' as a *Wnt11* gene, refuting secondary loss of this subfamily. Given that *Apis mellifera* (Hymenoptera) is less divergent within insects than *Drosophila* and *Anopheles*, and also that Ecdysozoa comprise Arthropods + Nematodes + Nematomorphs + Scalidophera¹⁶, this instance highlights the importance of wide taxonomic and genomic sampling when reconstructing phylogeny of large gene families (Brusca and Brusca 1990; Moore 2001).

ii) *Tu-WntC* affinity

Incomplete resolution by Bayesian inference

I used Bayesian inference to analyse 309bp *Tu-WntC* coding nucleotides, including 41 Wnt sequences from a wide range of other proto- and deuterostome taxa. Relationships between Wnt subfamilies in the consensus tree (Figure 4.1.2d) are largely unresolved, although *Wnt10*, *Wnt1*, *Wnt7* and *Wnt11* clades receive high (1.00 – 0.88) posterior probability support, and *Wnt3* and *Wnt4* clades branch with notable support (0.69 and 0.72 respectively). *Tu-WntC* affinity is unresolved within a reasonably stable (0.76) clade that includes all but *Wnt3* and *Wnt10* genes, indicating *Tu-WntC* orthology to almost any *Wnt* gene except *Wnt3* and *Wnt10*.

Wnt3, *Wnt10*, *Wnt9*, *Wnt1* + *Wnt6* have been proposed as members of an ancestral *Wnt* cluster, evolved prior to divergence of Cnidaria from other Eumetazoa (Kusserov et al. 2005; Prud'homme et al. 2002). Figure 4.1.2d branching suggests that *Wnt3* is most ancestral and *Wnt10* has a intermediate level of molecular divergence relative to remaining *Wnt* genes. This scenario could be explained by tandem duplication of *Wnt3* to give a proto-*Wnt10*, followed by further duplications to form subsequent proto-*Wnt* sub-family members.

Wnt gene loss in the Ecdysozoa

Little solid information can be gleaned regarding *Wnt* subfamily member loss, as resolution is either incomplete or not well supported at relevant nodes. *Caenorhabditis elegans*' *Cwn1* is identified as a *Wnt4* ortholog, supporting Prud'homme et al. (2002) but refuting Kusserov et al. (2005) whose phylogeny proposes a *Wnt1* affinity (Kusserov et al. 2005; Prud'homme et al. 2002). *Cwn2* is identified as a *Wnt8/Wnt9-like* ortholog (0.91) within a *Wnt8/9+Wnt5* clade (0.66), concurrent with *Wnt8* affinity proposed by previous authors.

¹⁶ Scalidophera = Priapulida + Kinorhyncha + Loricifera

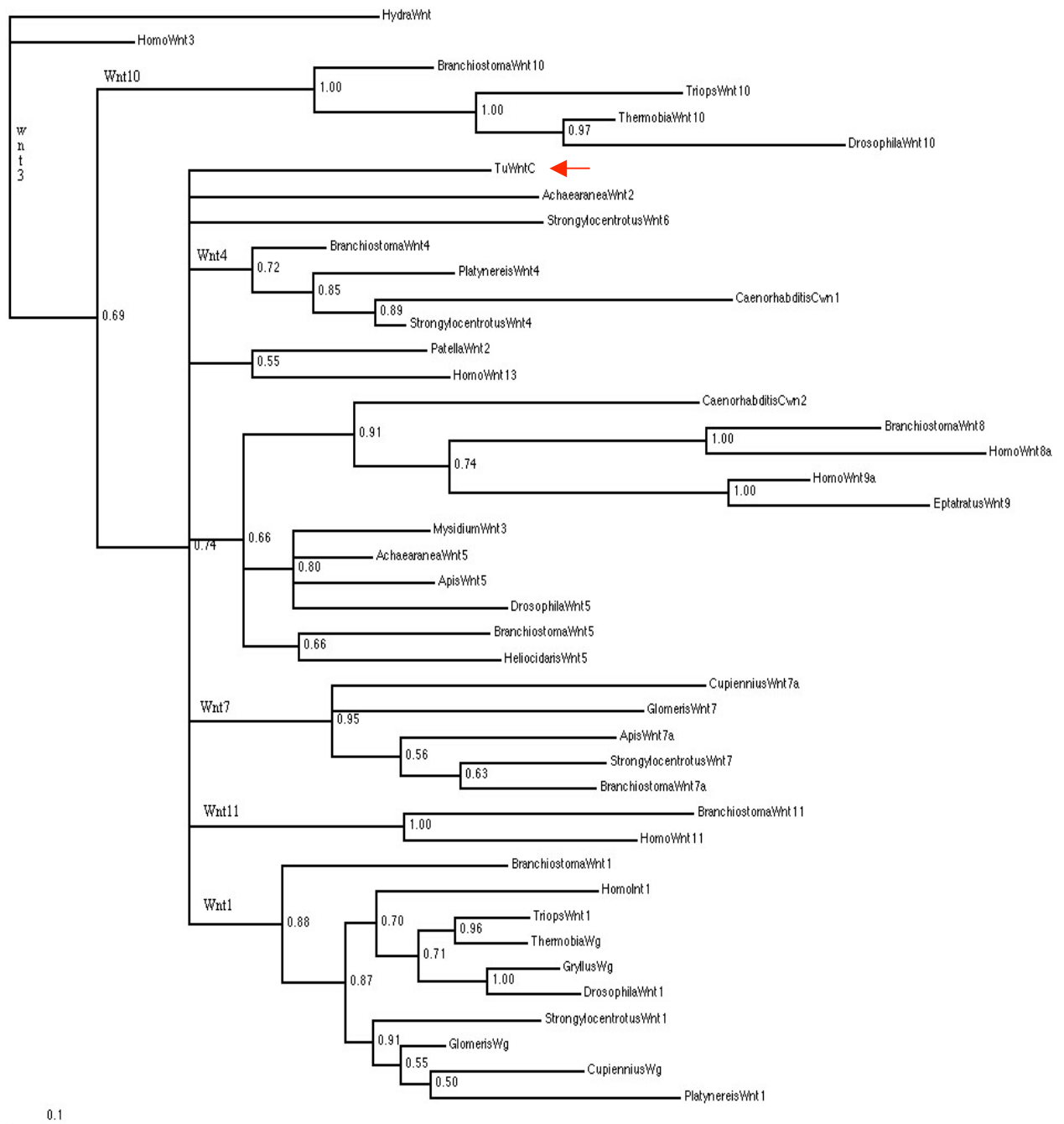


Figure 4.1.2d Bayesian inference consensus tree for analysis of *Tetranychus urticae* *WntC* coding sequence (258bp aligned CDS character set), with 42 Wnt sub-family genes from a range of proto- and deuterostome taxa. The affinity of *Tu-WntC* (red arrow) within the Wnt family is not resolved, although it is unlikely to belong to the better supported clades: *Wnt1*, *Wnt5*, *Wnt4*, *Wnt7*, *Wnt10* and *Wnt11*. *Wnt5* and *Wnt10* genes appear most ancestral within the *Wnt* superfamily (0.69/1.00 p.p.), in agreement with previous phylogenies.

Combined with information about *Apis Wnt2* as a possible *Wnt11* ortholog (Figure 4.1.2c), the *Tu-Wnt* analyses indicate a larger complement of conserved *Wnt* subfamilies in the Ecdysozoa than previously recognised, perhaps including *Wnt11* (e.g. *Apis*), *Wnt4* and a divergent *Wnt 5/8/9*-type gene (e.g. *C. elegans*).

4.2 Detecting Wnt1 protein activity in *Tetranychus urticae* with Wg and Armadillo antibodies

4.2.1 The Wnt signal pathway: antibody targets

An appreciation of Wnt signalling pathway components and interactions is instructive when attempting to interpret protein expression detected by antibodies against such components (c.f. Chapter I, Figure 1.2.1b). The Wnt ligand is targeted for secretion, upon which it binds the extracellular domain of a 7-pass trans-membrane Frizzled receptor protein to activate Dishevelled (Dsh), which is associated with the receptor intracellularly (Cliffe et al. 2003; Wallingford and Habas 2005; Wodarz and Nusse 1998). This Wnt-induced interaction occurs only in the absence of antagonistic ligand binding by sFRPs (secreted Frizzled-related proteins) or *Wnt* down-regulation by factors such as Nkd, Wf, Nemo, or Toll/NFkB signalling (Gerlitz and Basler 2002; Gordon and Nusse 2006; Jia et al. 2002; Jones and Jomary 2002; Moon et al. 1997; Rattner et al. 1997; Rousset et al. 2001; Simcox et al. 1989; Zeng et al. 2000; Zeng and Verheyen 2004). Dsh activation causes cytoplasmic Axin to localise at the cell membrane, and inhibits glycogen synthase kinase-3 (GSK-3 β) to prevent N' phosphorylation of Armadillo (Arm) bound within an Axin·Kinase·APC tumor suppressor complex (Amit et al. 2002; Benez 2003; Cliffe et al. 2003; Hayward et al. 2005; Johnson et al. 1993; Peifer et al. 1994; Peifer and Wieschaus 1990; Riggelman et al. 1990; Weitzel et al. 2004). Recent work suggests that downstream of Dsh, a PP2A (protein phosphatase sub-unit 2A) complex stabilises Arm, and that GSK-3-mediated phosphorylation critically stabilises Axin (Bajpal et al. 2004; Tolwinski and Wieschaus 2004; Yang et al. 2003). Armadillo is homologous to deuterostome β -catenin, and is a critical player in the Wnt pathway, transducing signal to the nucleus and activating target gene transcription. (Cliffe et al. 2003; Peifer and Wieschaus 1990). In the absence of Wnt activation, Arm is phosphorylated (at ser 45) and targeted for ubiquitin pathway degradation, but

GSK-3 β inhibition allows Arm to transfer to the nucleus, associate with a LEF family protein (e.g. TCF/Pangolin) and activate target transcription *via* LEF-mediated DNA binding (Amit et al. 2002; Gilbert 2000). For a recent review of Wnt signalling see Gordon & Nusse (2006) (Gordon and Nusse 2006).

As described in the following paragraphs, I tested antibodies against the Wg/Wnt1 ligand at the top of the signalling pathway, and against Arm/ β -catenin. Arm function as transducer and transcriptional activator is known to depend on Wnt activation, and hence I aimed to look at Arm protein expression as it accurately mirrors, and serves as a proxy to reveal, Wnt ligand distribution (Brook and Cohen 1996; Dhawan and Gopinathan 2003; Riggleman et al. 1990).

4.2.2 Wingless monoclonal antibody 4D4

Brook & Cohen (1996) used an immunogen comprising amino acids 3 - 468 of *Drosophila melanogaster* Wingless protein to immunise a mouse donor, and the resulting monoclonal antibody 4D4 was shown to be specific for, and reveal localisation of, *Drosophila melanogaster* Wingless (Brook and Cohen 1996). Whole mount immuno-staining *Tetranychus urticae* embryos with Wingless 4D4 antibody, however, was unsuccessful: I detected no protein expression by means of DIG-AP chromogenesis (c.f. section 4.2.4). As the 4D4 antibody does not report hypothetical Wg/Wnt1 protein localisation, its detection cannot serve as a proxy for *Wnt1* mRNA expression patterns; those being unobtainable due to lack of a cloned *Tu-Wg* ortholog.

Cross-reactivity with the α -Wingless 4D4 antibody has been reported in *Bombyx mori*, but as a Lepidopteran, this species is separated from Diptera by a lesser divergence time relative to *Tetranychus* – i.e. approximately 250Ma rather than ~540Ma (Dhawan and Gopinathan 2003). I propose that the *Drosophila* epitope targeted by α -Wingless 4D4 is sufficiently modified in *Tetranychus* to abrogate antibody binding, largely due to the evolutionary separation and divergence time between dipteran and acarid lineages.

4.2.3 Armadillo/ β -Catenin antibodies

Antibodies have been raised against two slightly overlapping domains of the *Drosophila melanogaster* **Armadillo** protein; antibody N1 recognises a 47 amino acid epitope (*arm* cDNA bp 1880-2110) and N2 recognises a 58 amino acid epitope (bp 2011-2175). Polyclonal α -Arm N1 and N2 antibodies were generated by rabbit immunisation, and a monoclonal α -Arm N2 antibody from a mouse

donor(Riggleman et al. 1990). Rabbit polyclonal and mouse monoclonal antibodies have also been raised against un-phosphorylated **β -catenin**, the vertebrate Armadillo homolog; in both cases the short *Homo* peptide chosen was highly conserved and therefore expected to produce a broadly cross-reactive antibody (Johnson et al. 1993; Takeichi 1991). A rabbit polyclonal against *Homo* β -catenin aa 768-781, and a mouse monoclonal (8E4) against *Homo* β -catenin aa 27-37 are available from Sigma® and Upstate® respectively.

i) Arm N2 monoclonal antibody

Whole mount immuno-staining of *Tetranychus urticae* embryos with α -Arm N2 (monoclonal N2 7A1) available from DSHB (Developmental Studies Hybridoma Bank, Iowa) was unsuccessful; the antibody failed to cross-react with the target epitope of any hypothetical *Tetranychus urticae* Armadillo ortholog. Cross-reaction of α -Arm N2 outside *Drosophila* is known for *Manduca sexta*, a Lepidopteran species that, as regarding α -Wg in *Bombyx mori*, is but little removed from Diptera in evolutionary time compared to *Tetranychus* (250Ma vs. 540Ma) (Kraft and Jäckle 1994).

Conservation of Wnt signal pathway components in diverse metazoans, and the identification of at least three *Tu-Wnt* genes by preliminary EST screening of the *Tetranychus urticae* cDNA library leads me to infer that a spider mite Arm ortholog must exist (Gordon and Nusse 2006; Hobmayer et al. 2000; Imai 2003; Wikramanayake et al. 2003). Presumably, molecular evolution since the divergence of *Drosophila* (Hexapoda: Diptera) and *Tetranychus* (Chelicerata: Acarida) lineages has altered *Tu*-Arm epitope residues, making recognition by the *Drosophila*-specific N2 antibody ineffective.

ii) β -Catenin poly- and monoclonal antibodies

I demonstrated a lack of cross-reactivity in *Tetranychus urticae* embryos during immuno-staining experiments with Sigma® β -catenin polyclonal and Upstate® monoclonal 8E4 antibodies (primary antibody 1:10 or 1:100). I observed only background staining; no potential Wnt1 protein expression domains (c.f. section 4.2.4) were evident at any stage of embryogenesis.

It has since been reported that an antibody raised against *Xenopus* β -catenin cross-reacts in *Tetranychus urticae*, giving a near-ubiquitous staining pattern throughout the embryo(Khila 2006a). This uniform pattern could reflect:

- i) generalised, non-specific binding that generates an artefact, or
- ii) deployment of an Arm homolog in processes and pathways beyond those activated by Wnts in typical segmental and appendage domains, or

- iii) ubiquitous Arm translation, with specific activation controlled by some kind(s) of post-translational modification.

If the uninformative, ubiquitous pattern revealed with anti-*Xenopus* β -catenin represents 'real' ubiquitous protein expression, then the ubiquitous 'signal' ignored as background staining when I tested poly- and monoclonals against *Homo* β -catenin may also reflect genuine cross-reactivity and Arm protein. However, whether or not any of these uniform patterns prove genuine, a *Tetranychus* Arm/ β -catenin ortholog activated by multiple Wnt factors, or outside the canonical Wnt pathway, can hardly provide a useful means to detect specific Wnt1/Wg signal transduction.

4.2.4 Conserved arthropod Wnt1 roles: *Tetranychus* predictions?

Posterior parasegments

In a wide range of arthropods, early *wg* transcription occurs in segmental stripes - just anterior to Engrailed-positive cells and thus correlated with posterior parasegment boundaries (Damen 2002; Hejnol and Scholtz 2004; Hughes and Kaufman 2002a; Kraft and Jäckle 1994; Nulsen and Nagy 1999; Prpic 2004a; Prpic 2004b) reviewed in (Cornec and Gilles 2006; Davis and Patel 2002; Peel et al. 2005; Tautz 2004). The conserved function of *wingless* in establishing and maintaining parasegmental boundaries has been confirmed by functional studies in *Drosophila*, *Tribolium castaneum*, *Gryllus bimaculatus* and *Oncopeltus fasciatus*. However a later role for Wnt signalling in *Drosophila*, mediating proximo-distal patterning in ventral limbs, appears not to be conserved, even in more basally branching insects such as the hemipteran *Oncopeltus* and orthopteran *Gryllus bimaculatus* (Akam 1987; Angelini and Kaufman 2005a; Deutsch 2004; Miyawaki et al. 2004).

Wnt sub-/para-functionalisation

In *Cupiennius salei* and *Triops longicaudatus* segmentation, and in *Drosophila* tracheal development, the expression of multiple *Wnt* orthologs is required to mimic a function being executed by a single ortholog (*Wnt1* in the case of segmentation) in other taxa (Damen 2002; Nulsen and Nagy 1999; Llimargas & Lawrence, 2001).

Wnt conservation and Tetranychus

The prevalence of *Wnt* super-family genes and generic nature of Wnt signalling in mediating metazoan cell-cell interactions is striking: Wnt-related signals can initiate cellular/embryonic asymmetry, prescribe segments, segmental structures (including limbs and neuroanatomy) and guide diverse morphogenetic processes (Angelini and Kaufman 2005a; Cohen 1990; Couso and Martinez Arias 1994; Deutsch 2004; Janssen et al. 2004; McMahon and McMahon 1989; Momose and Houliston 2007; Peel 2004; Seaver and Kaneshige 2006; Weitzel et al. 2004; Wodarz and Nusse

1998). Therefore I would expect multiple *Wnt* orthologs to exist in *Tetranychus*, functioning in a range of segmental or non-segmental contexts. I predict that Wnt signalling in *Tetranychus* would be mediated parasegmentally by a *Wnt1* ortholog, or combination of orthologs whose sub-functionalisation mimics typical *Wnt1* deployment. For a more detailed discussion relating to *Wg/Wnt* and *Tetranychus* see the Discussion in Chapter VII.

4.3 Cloning *Tetranychus engrailed* (*Tu-en*) genes

4.3.1 Degenerate PCR and sequence analysis: Two *Tu-en* genes

i) Degenerate PCR products

Engrailed family proteins share 5 regions of homology, termed Engrailed homology (EH) domains EH1 – EH5. EH1 and EH2 are binding sites for Groucho and Extradenticle co-repressors respectively, EH3 links the EH2 domain with the conserved homeodomain EH4, and EH5 is involved in repression (Peel et al. 2006; Peltenburg and Murre 1996; Tolkunova et al. 1998). I designed degenerate primers to target conserved EH1, EH2 and EH5 domain sequences; EH2-Fwd and EH5-Rev were kindly donated by A. Peel (see Appendix 2.4 for complete primer sequences). With primers EH2-Fwd (GG CCC GCC TGG GTN TAY TG) and EH5-Rev (TG TGA GTG GTT GTA CAG GCC YTG NGC C), I amplified putative *Tetranychus urticae engrailed* (*Tu-en*) fragments from embryonic cDNA and 3 independent genomic DNA extracts. A touchdown PCR program was used to optimise *en*-specific binding, as explained with regard to previous gene amplification (program details in Materials & Methods, section 8.2).

Fragments obtained from cDNA were ~350bp in size, and those from gDNA either ~350bp or ~470bp, suggesting the presence of distinct copies of *engrailed*, one of which contains intronic DNA, removed in mature mRNA by post-transcriptional modification. Multiple clones of putative *Tu-en* PCR products were identified by BLAST searches as arthropod *engrailed* orthologs, and sequence analysis showed that two versions of the gene were indeed present; I named these provisional paralogs *Tu-en1* and *Tu-en2* (Figure 4.3.1a,i-iii). A comparative *Tu-En1/En2* protein alignment

(i) *Tu-en1* (cDNA)

```
1/1                                     31/11
CCC GCC TGG GTA TAC TGT ACT CGA TAC TCG GAT AGA CCG TCT TCA GGT CCT AGA TCA AGA
pro ala trp val tyr cys thr arg tyr ser asp arg pro ser ser gl pro arg ser arg
61/21                                     91/31
AAG TTA AGA AAA AAG GAG AAA AAG TCT GAT GAG AAA AGG CCT CGA ACC GCC TTT ACC GCC
lys leu arg lys lys glu lys lys ser asp glu lys arg pro arg thr ala phe thr ala
121/41                                     151/51
GAA CAA TTG CAA CGA TTG AAA CAA GAG TTT CAA GAT AAT AGA TAT TTA ACC GAG AAA CGT
glu gln leu gln arg leu lys gln glu phe gln asp asn arg tyr leu thr glu lys arg
181/61                                     211/71
CGT CAA GAT TTG GCT CGA GAT TTA AAG TTG AAC GAG TCA CAA ATT AAG ATT TGG TTC CAG
arg gln asp leu ala arg asp leu lys leu asn glu ser gln ile lys ile trp phe gln
241/81                                     271/91
AAT AAA CGG GCA AAG ATT AAG AAG GCC GGT GGT CAA CGG AAT CCA TTG GCT TAC CAT TTA
asn lys arg ala lys ile lys lys ala gly gly gln arg asn pro leu ala tyr his leu
301/101
ATG GCC CAG GGC CTG TAC AAC CAC TCC
met ala gln gly leu tyr asn his ser
```

(ii) *Tu-en1* (gDNA)

```
1/1                                     31/11
CCC GCC TGG GTA TAC TGT ACT CGA TAC TCG GAT AGA CCG TCT GCA GGT AAGTGTAGAGATTTT
pro ala trp val tyr cys thr arg tyr ser asp arg pro ser ala gl
131/17
TTCTCACTTCATTTTTGTGATTCTCTTTTGTTCATTCTGTATTCATCATTACCTTAATCAGGT CCT AGA TCG
pro arg ser
151/23                                     181/33
AGA AAG CTA AGA AAA AAG GAG AAA AAG TCT GAT GAG AAA AGA CCT CGA ACT GAC TTT ACC
arg lys leu arg lys lys glu lys lys ser asp glu lys arg pro arg thr asp phe thr
211/43                                     241/53
GCC GAA CAA TTG CAG CGA TTG AAA CAA GAG TTT CAA GAT AAT AGG TGA GTTTC AATGTTTGCT
ala glu gln leu gln arg leu lys gln glu phe gln asp asn arg -X-
325/55
AATTTGATGCGAGAAAAGTTTACTCATTTGTTGGAGTTTTCATCTTTTTTACATGTTTCAGA TAT TTA ACC GAG
tyr leu thr glu
340/60                                     370/70
AAA CGT CGT CAA GAT TTG GCT CGA GAT TTA AAG TTG AAC GAG TCA CAA ATT AAG ATT TGG
lys arg arg gln asp leu ala arg asp leu lys leu asn glu ser gln ile lys ile trp
400/80                                     430/90
TTC CAG AAT AAG CGG GCC AAG ATT AAG AAG GCC GGT GGT CAA CGG AAT CCA TTG GCT TAC
phe gln asn lys arg ala lys ile lys lys ala gly gly gln arg asn pro leu ala tyr
460/100
CAT TTA ATG GCC CAA GGC CTG TAC AAC CAC TCC
his leu met ala gln gly leu tyr asn his ser
```

(iii) *Tu-en2* (gDNA)

```

1/1                               31/11
CCC GCC TGG GTG TAC TGC ACC CGC TAC TCC GAC CGA CCG TCT TCA GGT CCT CGA TCT CGC
pro ala trp val tyr cys thr arg tyr ser asp arg pro ser ser gly pro arg ser arg
61/21                               91/31
AAA ATC AAA AAG AAG GAG AAG AAG GAC GAG AAA AGG CCG CGC ACA GCA TTC ACC GCA GAG
lys ile lys lys lys glu lys lys asp glu lys arg pro arg thr ala phe thr ala glu
121/41                               151/51
CAA CTC GCT CGA CTA AAG CAG GAA TTC ACG GAA AAT AGA TAT CTT AAT GAA AAA CGG CGA
gln leu ala arg leu lys gln glu phe thr glu asn arg tyr leu asn glu lys arg arg
181/61                               211/71
CAA GAC CTT GCT CGC GAA TTA AAA CTG AAC GAG TCG CAA ATC AAA ATT TGG TTT CAA AAC
gln asp leu ala arg glu leu lys leu asn glu ser gln ile lys ile trp phe gln asn
241/81                               271/91
AAG CGA GCC AAA ATA AAG AAA ACA ACC AGT ACG CGC AAC CCA TTG GCA TTG CAC TTG ATG
lys arg ala lys ile lys lys thr thr ser thr arg asn pro leu ala leu his leu met
301/101
GCC CAA GGC CTG TAC
ala gln gly leu tyr

```

(iv)

	1	11	21	31	41	51	
En-1	PAWVYCTRYS	DRPSSGPRSR	KLRKKEK	KSD	EKRPRTAFTA	EQLQRLKQEF	QDNRYLTEKR
En-2	PAWVYCTRYS	DRPSSGPRSR	KIKKKEK	K-D	EKRPRTAFTA	EQLARLKQEF	TENRYLNEKR
			**	*		*	** *

	60	70	80	90	100
En-1	RQDLARDLKL	NESQIKIWFQ	NKRAKIKKAG	GQRNPLAYHL	MAQGLY
En-2	RQDLARELKL	NESQIKIWFQ	NKRAKIKKTT	STRNPLALHL	MAQGLY
	*		** **	*	

Figure 4.3.1a Nucleotide and amino acid sequence data for *Tetranychus urticae engrailed* genes retrieved by degenerate PCR. (i) *Tu-en1* data, from cDNA. (ii) *Tu-en1* data including introns, obtained from genomic DNA template. (iii) *Tu-en2* data, obtained from genomic DNA. Boxed: EH2-Fwd and EH5-Rev primer target sequences. (iv) Comparison of Tu-En1 and -En2 proteins, revealing 87.7% shared identity. Asterisk: divergent amino acid (13 of 106 residues; 12.3%). Red font: serine insertion → En1 KSD motif vs. En2 KD motif. Green: EH2/Exd binding domain. Blue: EH3 domain. Grey: EH4/Homeodomain. Yellow: EH5 repression domain [Damen, 2002; Peel et al, 2006].

indicates 87.7% shared identity at the amino acid level within the conserved EH2-EH5 region represented (Figure 4.3.1a,iv).

ii) Prevalence of *engrailed* duplication

Duplications of the *engrailed* gene are evident within various metazoan lineages, including myriapods, chelicerates, crustaceans, hexapods, cephalopod molluscs and vertebrates (Bastianello, 2001; Damen, 2002; Wray, 1995; Peterson, 1998; Abzhanov, 2000; Quéinnec, 1999; Gibert, 2000; Peel, 2006). Single *engrailed* orthologs are present in echinoderms, urochordates and cephalochordates but two (or more) *engrailed* genes, each on separate chromosomes, are present in craniates, indicating independent duplication related to whole genome duplication in the vertebrate lineage (Holland and Williams 1990; Jiang and Smith 2002; Kobayashi et al. 2003; Lowe et al. 2003; Rida et al. 2004). Among lophotrochozoans, platyhelminthes, annelids and almost all molluscs possess one *engrailed* gene; the only duplication so far reported is in cephalopod *Nautilus pompilius*, probably representing independent tandem duplication within this specific mollusc lineage (Bely and Wray 2001; Prud'homme et al. 2003; Seaver and Kaneshige 2006; Shankland and Seaver 2000; Wedeen and Weisblat 1991; Wray et al. 1995). In the absence of genomic data for the lophotrochozoan species examined, however, the possibility of undetected duplicates or basal duplication followed by extensive gene loss, cannot be ruled out.

Only a single *engrailed* gene was found in the acarid *Archegozetes longesitosus* and the spider *Achaeearanea tepidariorum*, but a paralogous pair has been reported for the spider *Cupiennius salei* as well as here in *Tetranychus urticae* (Akiyama-Oda and Oda 2006; Damen 2002; Telford and Thomas 1998a). Evidence of mite and spider species with both one (*Archegozetes*, *Achaeearanea*) and two (*Tetranychus*, *Cupiennius*) distinct *engrailed* genes indicates either:

- i) independent duplications in Arachnid and Acarid lineages, or
- ii) earlier duplication followed by independent gene loss –

unless second copies of *engrailed* exist undetected for *Archegozetes* and *Achaeearanea*, which is not testable due to lack of whole genome data.

iii) An EH2 C-terminal RX motif

RX motif distribution in Eumetazoans An RX motif located directly after ser-ser-gly of the EH2 domain C terminus is noticeable in all hexapods and some crustaceans (*Artemia*) as a conserved RS motif, in some myriapods as RS (*Glomeris*) or RR (*Lithobius*), in the annelid *Platynereis* as RR and in the cephalochordate *Branchiostoma* as RT (Figure 4.3.2b). No such motifs have been observed thus

far in molluscs or chelicerates, and the *Tetranychus* En1 and En2 sequences prove no exception (Figure 4.3.1a,iv).

Functional significance RX motifs are implicated in satisfying binding requirements for phosphorylation by Ca²⁺ dependent kinases, augmenting the potential for targeted phosphorylation within surrounding serine-rich domains (Krebs and Beavo 1979; Peel et al. 2006). The wide taxonomic spread of SSG-RX motifs could indicate convergent insertion of RX within EH2 domains of various metazoan lineages, perhaps modulating levels of phosphorylation and its effects on Engrailed localisation or activity.

Evolutionary interpretation of arthropod RX motif distribution Focussing on arthropods, a lack of RX motif in chelicerates, but presence in some myriapods (RS/RR), most crustaceans (RS) and all hexapods (RS) is interpreted differently depending on the phylogenetic placement of myriapods:

A. If chelicerates are basal to the mandibulate phyla (myriapods + pancrustaceans), as concluded by (Garey et al. 1996; Giribet 2002; Giribet et al. 1996; Giribet et al. 1999; Hughes and Kaufman 2001) and (Blaxter 2001), three possible scenarios emerge:

- i) *engrailed* duplication before major arthropod phyla diverged, and RX motif insertion into one paralog at the base of the mandibulates;
- ii) *engrailed* duplication at the base of the arthropods, independent insertion of RX motifs in some myriapod lineages, and insertion of an RS motif at the base of the Pancrustacea (insects + crustaceans);
- iii) *engrailed* duplication at the base of the arthropods, independent insertion of RX motifs in some myriapod lineages, some crustaceans lineages, and insertion of an RS motif at the base of the insects.

B. If chelicerates and myriapods are sister groups, as reported in molecular phylogenies of (Cook et al. 2001; Giribet et al. 1996; Hwang et al. 2001; Kusche and Burmester 2001) and (Blaxter 2001) possible scenarios include:

- i) *engrailed* duplication + RX insertion at the base of the arthropods, then RX motif loss from chelicerates, some myriapods and some crustacean lineages;
- ii) independent *engrailed* duplications within some chelicerate and myriapod lineages, and duplication + RX motif insertion at the base of the Pancrustacea, followed by independent gain of RX motifs in some myriapods and loss of RX motifs in some crustaceans!

I consider **A(i)** to be the most parsimonious scenario, favouring the basal position of chelicerates relative to myriapods (c.f. Chapter I, section 1.1.3). A conclusive phylogeny for chelicerates and myriapods in relation to Pancrustacea, as well as comprehensive (i.e. genomic) data on *engrailed* paralogs and signature motifs, would help clarify the true pattern of *engrailed* gene evolution in arthropods. More complete coverage of arachnid taxa (including scorpions and lesser known

groups), Xiphosura, myriapods and confirmation of a single *en* gene (with RS motif) in *Artemia* would be particularly useful.

4.3.2 Inverse PCR (iPCR): Further *en-1* and *en-2* sequence data

i) Experimental conditions

I carried out inverse PCR using 3 independent genomic DNA extracts, each split into 4 aliquots and digested with restriction enzymes NotI, SalI, EcoRV and XhoI prior to circularisation (see Materials & Methods, section 8.2.3 for details). Multiple PCR conditions were tested, annealing at either 50°C or 49°C and with 3.0mM, 3.2mM or 3.4mM [Mg²⁺]. A lower annealing temperature facilitates primer binding and elevated magnesium concentrations have a similar effect, large Mg²⁺ ions theoretically inserting within the DNA helix, forcing the structure open for access to oligonucleotides. I designed iPCR primers facing outwards from known *Tu-en1* and *Tu-en2* sequence obtained by degenerate PCR (full primer details in Appendix 2.4). Putative *engrailed* fragments between ~300bp and 1Kb in size were successfully amplified, primarily for *Tu-en1* but some shorter fragments pertaining to *Tu-en2*. Regarding successful iPCR primer pair combinations, iEn1-L131+U173 with nested iEn1-L21+U240 gave *Tu-en1* products, and iEn2-L108+U208 with nested iEn2-L40+U241 amplified *Tu-en2* products. A semi-nested second iPCR reaction also gave possible products when priming with iEn1-L108+U241 after initial L40+U241, and iEn2-L21+U240 after initial L21+U173. Organisation of resulting sequence information into genomic orientation (c.f. Chapter VIII: Material & Methods, Figure 8.2.3), revealed almost 1.2Kb of *Tu-en1* and 360bp of *Tu-en2* (Figure 4.3.2a). The *Tu-en1* sequence included 381bp 5' UTR and 663bp CDS, interrupted by 2 small introns, 82bp and 80bp in size (Figure 4.3.2a, i & ii).

ii) Gene sequence analysis: EH domain coverage

Nucleotide and translation data for *Tu-en1* and *Tu-en2* sequences obtained by iPCR are given in Figure 4.3.2a, including protein codes, outermost primer sites and Engrailed-family homology (EH) domains where relevant. 714bp novel sequence was obtained for *Tu-en1*: 381bp 5' UTR followed by 309bp new coding sequence (including the Groucho binding domain EH1) and 24bp 3' CDS, coding for 8 further the EH5 domain amino acids (Figure 4.3.2a,i-ii). Minimal novel data was obtained for *Tu-en2*: 18bp 5' sequence coding for 6 amino acids, to complete and reach upstream of the EH2

(i)

```
1
ATGACATTTCCAACGTGTTTCAATTGATTAATCAATGAATGTAAATTAAGTTAGAAGCTAAT
62
TTCAGTTTTTGTCTCAAAAAATTAGTGTGGTGAACGGACTCTATAACGCAAAGATAGAGTTTTTCATAACAACAAAACCAA
142
CGTTCAAATGATCAAAACTCACTTTTATAAGACAAAAACTCTCGCGTTAAAGAGTTGCTCAGTATGGAGATAAGATCTA
222
TTGTAACTTTTGACTTTTGTGTTTTGAGCTCAAACCTCTGTTATAGTCTATGGTGAAGGTTGTCCTTTTGCCCTG
302
TTCCCTCTCATCCTCAACCTTACCTTACCTTTTTCAGTGTCTGTTACCTTGTTCGGTTCCTATTTAATCAAACAAAC
382/1
402/11
ATG GTT TCA TCG GTA TTC CCA CCC ACT AAC CCA ATC AAA AGT TTA AAA TTT TCA ATT GCC
Met val ser ser val phe pro pro thr asn pro ile lys ser leu lys phe ser ile ala
442/21
472/31
AGC ATT TTA TCA CCT GAA TTT TGT AAA TCG ACA ATT TTA AAT GAA ATA AAG ACA ACA GCC
ser ile leu ser pro glu phe cys lys ser thr ile leu asn glu ile lys thr thr ala
502/41
532/51
ACC GTC ATG AAT GTG AAC AAT GGA AGC AAA AAA CGG AGA CAC AAC GAA AGT GAA AGC TCT
thr val met asn val asn asn gly ser lys lys arg arg his asn glu ser glu ser ser
562/61
592/71
GAA AGT GAA ACA TGT GCA AGG AAA AAA GTT TCG TGT GAA ATA TCT GAT AGC GAA TCT TCA2
glu ser glu thr cys ala arg lys lys val ser cys glu ile ser asp ser glu ser ser
622/81
652/91
AAT AGT TCA TCA TTA AAC ACC CAA GAA GGA AAC AGT TCA TCT CTA CTA TCA TCA CCC AAC
asn ser ser ser leu asn thr gln glu gly asn ser ser ser leu leu ser ser pro asn
682/101
712/111
TCA TCC TTG CCC GCT TGG GTT TAC TGT ACT CGA TAC TCG GAT AGA CCG TCT GCA GGT
ser met leu pro ala trp val tyr cys thr arg tyr ser asp arg pro ser ala gln
729
AAGGTAGAGATTTTTTCTCACTTCATTTTTGTGATCTCTTTTTGTTTACATTCTGTATTTCATTCACCTTAATCA
812/120
842/130
GGT CCT AGA TCG AGA AAG CTA AGA AAA AAG GAG AAA AAG TCT GAT GAG AAA AGA CCT CGA
pro arg ser arg lys leu arg lys lys glu lys lys ser asp glu lys arg pro arg
869/139
899/149
ACT GAC TTT ACC GCC GAA CAA TTG CAG CGA TTG AAA CAA GAG TTT CAA GAT AAT AGG TGA
thr asp phe thr ala glu gln leu gln arg leu lys gln glu phe gln asp asn arg -X-
929
TGAGTTTCAATGTTTGCTAATTTGATGCGAGAAAAGTTTACTCATTTGTTGGAGTTTTCATCTTTTTTACATGTTTCA
1008/159
1038/169
TAT TTA ACC GAG AAA CGT CGT CAA GAT TTG GCT CGA GAT TTA AAG TTG AAC GAG TCA CAA
tyr leu thr glu lys arg arg gln asp leu ala arg asp leu lys leu asn glu ser gln
1068/179
1098/189
ATT AAG ATT TGG TTC CAG AAT AAG CGG GCC AAG ATT AAG AAG GCC GGT GGT CAA CGG AAT
ile lys ile trp phe gln asn lys arg ala lys ile lys lys ala gly gly gln arg asn
1128/199
1158/209
CCA TTG GCT TAC CAT TTA ATG GCC CAA GGC CTG TAC AAC CAC TCC ACT GCA TCA AAG CAA
pro leu ala tyr his leu met ala gln gly leu tyr asn his ser thr ala ser lys gln
1188/219
GGT GAC CCG
gly asp pro
```

(ii)

```
1
MVSSVFPPPTNPIKS LKFSIASILSPEFC KSTILNEIKTTATVMNVNNGSKRRRHNESESESETCARKKVSCEISDSESSN
82
SSSLNTQEGNSSLLSSPNS SLPAWVYCTRYSDRPSAG-intron1-PRSRKLRKKEKKSDEKRPRTDFTAEQLQRLKQEF
154
QDNRX-intron2-YLTEKRRQDLARDLKLNESQIKIWFQNKRAKIKKAGQQRNPLAYHLMAQGLYNHSTASKQGDP
```

Figure 4.3.2a

(iii)

```
1/1                               31/11
CCG CGG GAA TTC GAT TGG CCC GCC TGG GTA TAC TGC ACC CGC TAC TCC GAC CGA CCG TCT
pro arg glu phe asp trp pro ala trp val tyr cys thr arg tyr ser asp arg pro asp
61/21                               91/31
TCA GGT CCT CGA TCT CGC AAA ATC AAA AAG AAG GAG AAG AAG GAC GAG AAA AGG CCG CGC
ser ala pro arg ser arg lys ile lys lys lys glu lys lys asp glu lys arg pro arg
121/41                               151/51
ACA GCA TTC ACC GCA GAG CAA CTC GCT CGA CTA AAG CAG GAA TTC ACG GAA AAT AGA TAT
thr ala phe thr ala glu gln leu ala arg leu lys gln glu phe thr glu asn arg tyr
181/61                               211/71
CTT AAT GAA AAA CGG CGA CAA GAC CTT GCT CGC GAA TTA AAA CTG AAC GAG TCG CAA ATC
leu asn glu lys arg arg gln asp leu ala arg glu leu lys leu asn glu ser gln ile
241/81                               271/91
AAA ATT TGG TTT CAA AAC AAG CGA GCC AAA ATA AAG AAA ACA ACC AGT ACG CGC AAC CCA
lys ile trp phe gln asn lys arg ala lys ile lys lys thr thr ser thr arg asn pro
301/101                               331/111
TTG GCA TTG CAC TTG ATG GCA CAA GGC CTG TAC AAC CAC TCC ACA CTC CCG GCC GCC ATG
leu ala leu his leu met ala gln gly leu tyr asn his ser thr leu pro ala ala met
```

(iv)

```
1
PREFDWPAAVYCTRYSDRPSSGPRSRKIKKKEKKDEKRPRTAFTAEQLARLKQEFTEENRYLNEKRRQDLARELKLNESQIK
82
IWFQNKRAKIKKTTSTRNPLALHLMAQGLYNHSTLPAAAM
```

Figure 4.3.2a (i) Nucleotides (1196bp) and translated sequence data for *Tu-en1*, including 5' UTR and sequence coding for 221aa, interrupted by intron 1 = 82bp and intron 2 = 80bp. Boxed: outermost 'reverse' iPCR primer L21 at 5', 'forward' iPCR primer U240 at 3' end. (ii) Summary of Tu-En1 protein sequence (221 amino acids). (iii) Coding nucleotide sequence and translated data for Tu-En2 sequence, recovered by inverse PCR. Boxed: outermost 'reverse' iPCR primer L40 at 5' end, 'forward' iPCR primer U241 at 3' end. (iv) Summary of Tu-En2 protein sequence (120 amino acids). **M**: initiation codon (methionine). **X**: termination codon. Red: **EH1**/Groucho binding domain. Green: **EH2**/Exd binding domain. Blue: **EH3** linker domain. Grey: **EH4**/homeodomain (Tu-En1 63, Tu-En2 61 amino acids; final N' residues missing). Yellow: **EH5** domain residues within homeodomain region [Damen, 2002; Peel et al, 2006].

domain, and 18bp 3' coding sequence, adding 6 further amino acids to the EH5 domain sequence (Figure 4.3.2a,iii-iv).

I compared regions of *Tu-en1* and *Tu-en2* coding sequence that could be aligned with 32 *engrailed* orthologs from a range of arthropods (Ecdysozoa), annelids, molluscs (Lophotrochozoa), echinoderms, hemichordates and chordates (Deuterostomia). The corresponding multiple amino acid alignment (Figure 4.3.2b) encompasses 129 residues across domains EH1 to EH5.

4.3.3 Phylogenetic analysis of *Tu-en1* and *Tu-en2*

I applied Bayesian inference methods to the nucleotide alignment from which Figure 4.3.2b is derived, excluding 3rd codon base positions as usual due to an elevated tendency to transition and transversion which confounds consensus topologies (Baldauf 2003). The consensus tree topology (Figure 4.3.3) is not fully resolved, but *Tetranychus en1* and *en2* are clearly placed within an arthropod *engrailed* orthology clade.

Where present, *en* paralogs pair either according to species (e.g. *Drosophila*, *Bombyx*, *Periplaneta*, *Tetranychus*) or in unresolved, separate groups (e.g. *Sacculina* and *Procambarus* paralogs). Former, parallel pairings suggest closer sequence similarity between species paralogs than with any ortholog from another species, indicative of an independent and recent duplication that leaves little time for molecular divergence. Conversely, paralog separation above species level indicates an earlier duplication, followed by independent divergence in each gene copy. As the Bayesian analysis incorporated multiple EH domains, it may be that misleading effects were introduced due to concerted evolution, as discussed in Peel et al (2006) regarding domain conversion during evolution of hexapod *engrailed* paralogs (Peel et al. 2006). Domain conversion sustains regions of high sequence similarity between corresponding genes, reducing the degree of divergence observed between duplicates and thus creating the false impression of a recent, independent event.

Cases of species-specific paralogy pairings in the Figure 4.3.3 consensus may conceivably be used to support a model of i) multiple independent *engrailed* duplications or, given possible gene conversion as mentioned above, ii) a basal arthropod tandem duplication event followed by lineage specific insertions and gene losses. It is not possible without adding further data to determine which option, (i) or (ii), may be true.

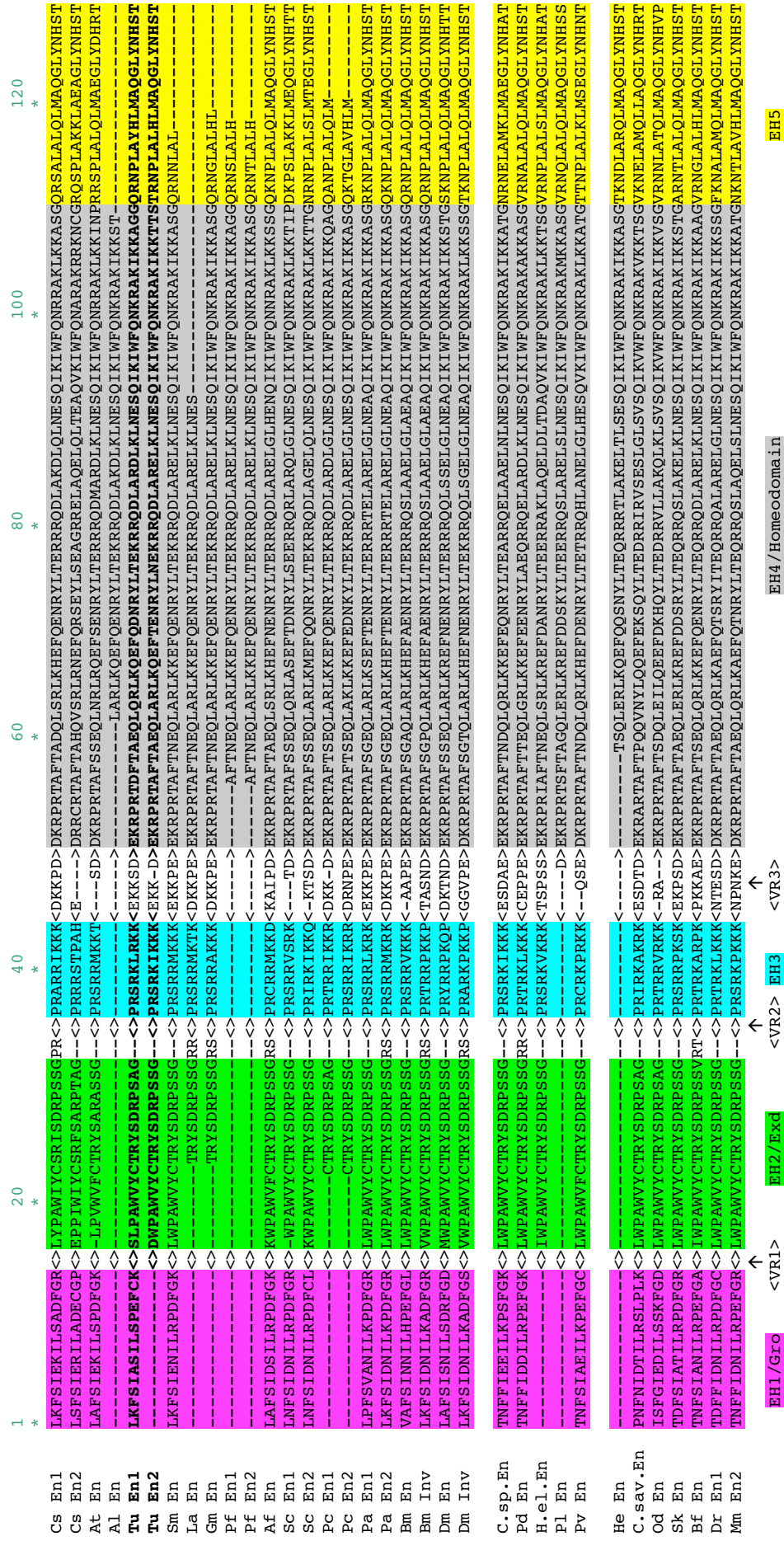


Figure 4.3.2b Multiple sequence alignment for 34x 129 Engrailed amino acids, including conserved Engrailed Homology domains EH1-EH5. Purple: EH1/Groucho cofactor binding domain. Green: EH2/Exd cofactor binding domain. Blue: EH3 linker/repressor domain. Grey: EH4/Homeodomain. Yellow: EH5 repression domain [Damen, 2002; Peel et al, 2006]. Bold: *Tetrahynchus urticae* En1 and En2 sequences. <VR>: Variable Region. VR1 <254aa; VR2 <17aa; VR3 <25aa - including small region of good homology (shown). Single-row gaps separate arthropods (top), lophotrochozoans - annelids+molluscs (middle) and deuterostomes (lower group).

Abbreviations: Cs - *Cupressus salei*; At - *Achaeaurea tepidariorum*; Al - *Archegonites longistylus*; Tu - *Tetranychus urticae*; Sm - *Strigamia maritima*; La - *Lalobius atkinsoni*; Gm - *Glomeris marginata*; Pf - *Pachmerium ferrugineum*; Af - *Artemia franciscana*; Sc - *Sacculina carcini*; Pc - *Sacculina americana*; Bm - *Bombyx mori*; Dm - *Drosophila melanogaster*; Csp - *Chaetopterus sp.*; Pd - *Platynereis dumerilii*; H.el. - *Hydroides elegans*; Pl - *Platella leidy*; Pv - *Patella vulgata*; He - *Helicodiaris erythrogramma*; C.sav. - *Ciona savignyi*; Od - *Oikopleura dioica*; Sk - *Saccoglossus kowalewki*; Bf - *Branchiostoma floridae*; Dr - *Danio rerio*; Mm - *Mus musculus*.

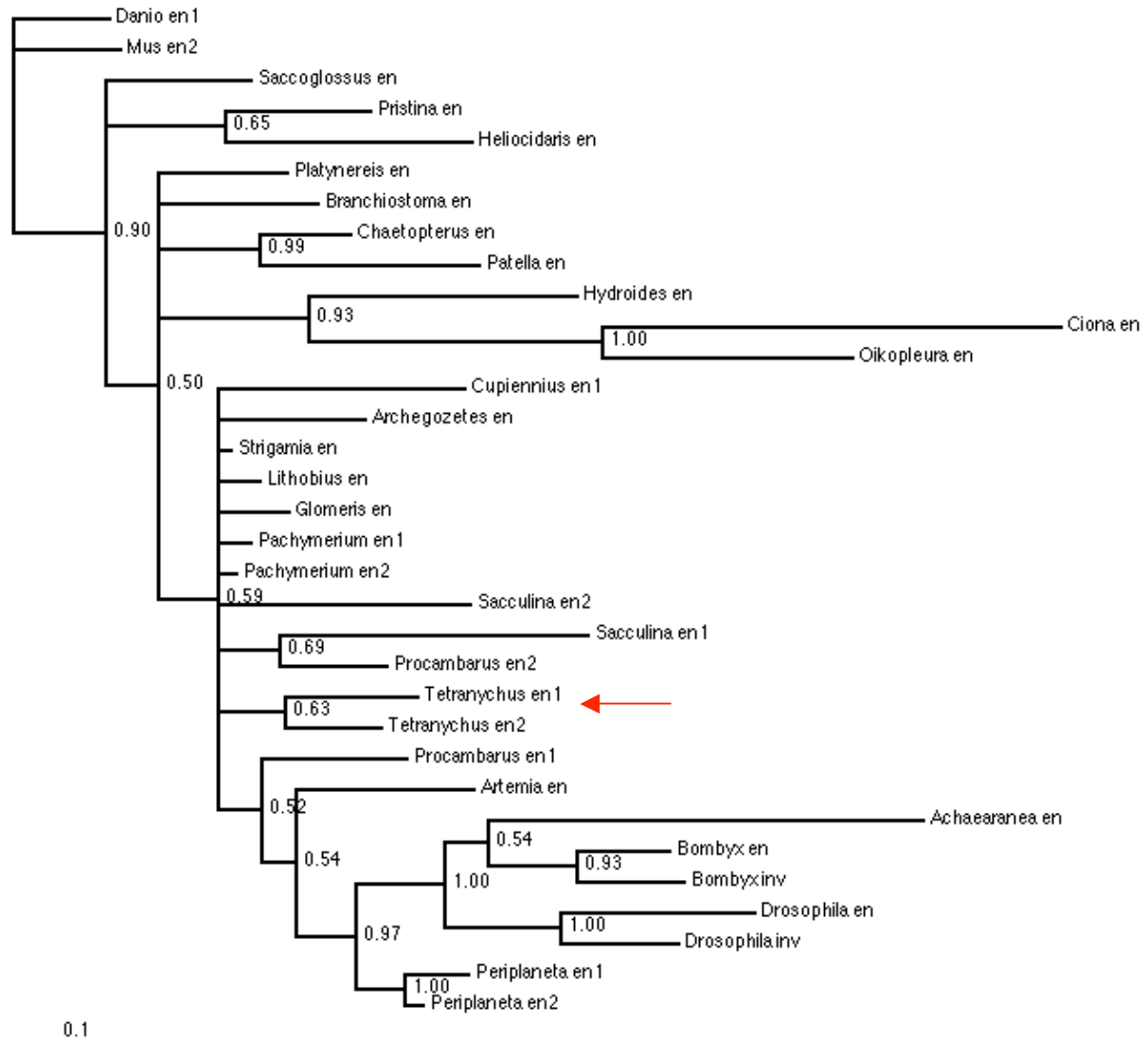


Figure 4.3.3 Bayesian consensus tree for analysis of 387bp *Tu-en1* and *Tu-en2* in relation to *en* ortho/paralogs from 25 other taxa: 13 arthropods, 5 lophotrochozoans and 7 deuterostomes. Although the topology is not fully resolved, *Tu-en* genes (red arrow) fall clearly within an arthropod *engrailed* orthology clad and branch together as non-identical paralogs.

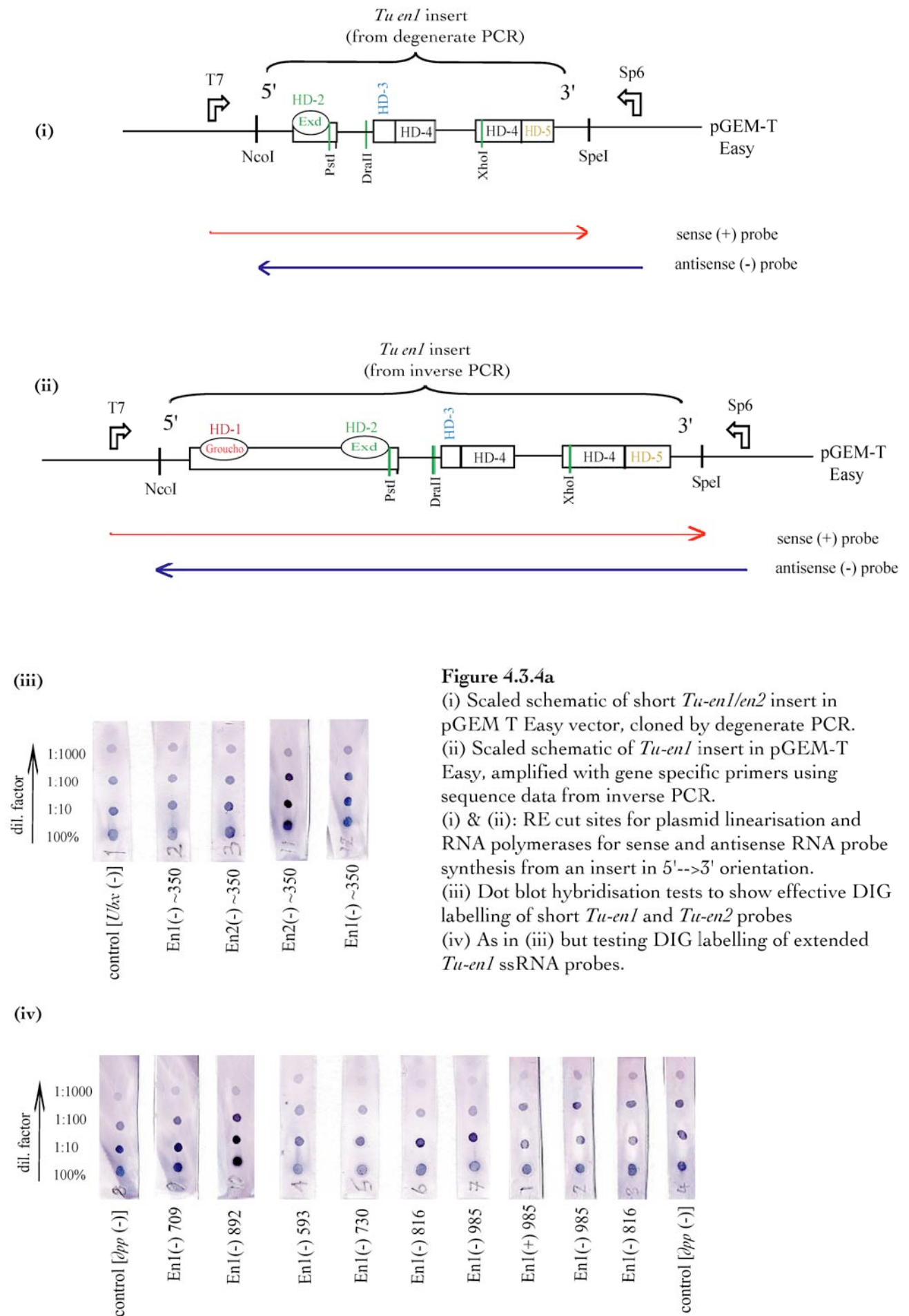
4.3.4 ssRNA *Tu-en* probe synthesis

i) Preliminary (short) *Tu-en1* and *Tu-en2* RNA probes

I synthesised single stranded DIG-labelled RNA probes from *Tu-en1* and *Tu-en2* inserts of ~350bp in length, cloned into pGEM®-T Easy vector following amplification by degenerate PCR. Samples of plasmid DNA were digested with EcoRI, cutting the vector at both 5' and 3' sites to excise and confirm both correct insert size and adequate DNA concentration. For ssRNA probe synthesis proper, restriction enzymes and RNA polymerases were selected depending on both insert and desired mRNA strand orientation: NcoI or SpeI for plasmid linearization and Sp6 or T7 RNA polymerase for transcription catalysis (Figure 4.3.4a,i). I verified effective α -DIG labelling of probe RNA by hybridising serially diluted samples of both sense, anti-sense and control probes with α -DIG-AP. Subsequent phosphatase reaction revealed less intense chromophore signal at higher dilution factors (Figure 4.3.4a, iii), indicating normal molar kinetics, successful DIG labelling and theoretical readiness for embryonic *in situ* hybridisation.

ii) Extended *Tu-en1* RNA probes

Having obtained gene sequence data for 1.2Kb *Tu-en1* by inverse PCR, I designed *Tu-en1*- specific primers and amplified several regions of the gene (from 5' UTR to EH5 domain) to provide raw material for RNA probes: primer target locations with respect to *Tu-en1* EH domain structure are shown in Figure 4.3.4b,i-ii. Touchdown PCR with 50°C final annealing temperature enabled amplification of *Tu-en1* fragments 593bp - 985bp in length from genomic DNA (Figure 4.3.4b,ii). I cloned *Tu-en1* products into pGEM®-T Easy vector, and carried out confirmation sequencing as well as sample EcoRI digestion to confirm expected insert size and adequate [DNA] for use in RNA probe synthesis (Figure 4.3.4b, iii). Depending on *Tu-en1* insert orientation in pGEM®-T Easy vector and whether I was synthesising sense or anti-sense RNA, plasmid linearization was mediated by NcoI or SpeI, and RNA transcription by Sp6 or T7 polymerase, (Figure 4.3.4a,ii). As previously, I tested sense, anti-sense and control mRNA probes for effective DIG-labelling by hybridisation with α -DIG-AP against serially diluted RNA probe samples bound to nitrocellulose membranes. All the results of such tests (Figure 4.3.4a,iv) verified successful DIG labelling and thus no theoretical barrier to detection of target-bound probes during embryonic *in situ* hybridisation.



4.4 *Tetranymphus engrailed* mRNA expression

4.4.1 *Tu-en* mRNA transcription in whole mount embryos

i) Negative results

I used single stranded, DIG-labelled *Tu-en1* RNA probes between 350bp and 985bp long, and ~350bp *Tu-en2* probes in whole mount *in situ* hybridisation of *Tetranymphus* embryos from blastoderm to germband contraction stage (9hr – 50hr AEL). No staining pattern was detected at any stage of embryogenesis (Figure 4.4.1). I observed elevated background signal ubiquitously in limb and cephalic tissues during germband contraction, presumably related to non-specific probe localisation within this region of dense tissue (Figure 4.4.1,vi).

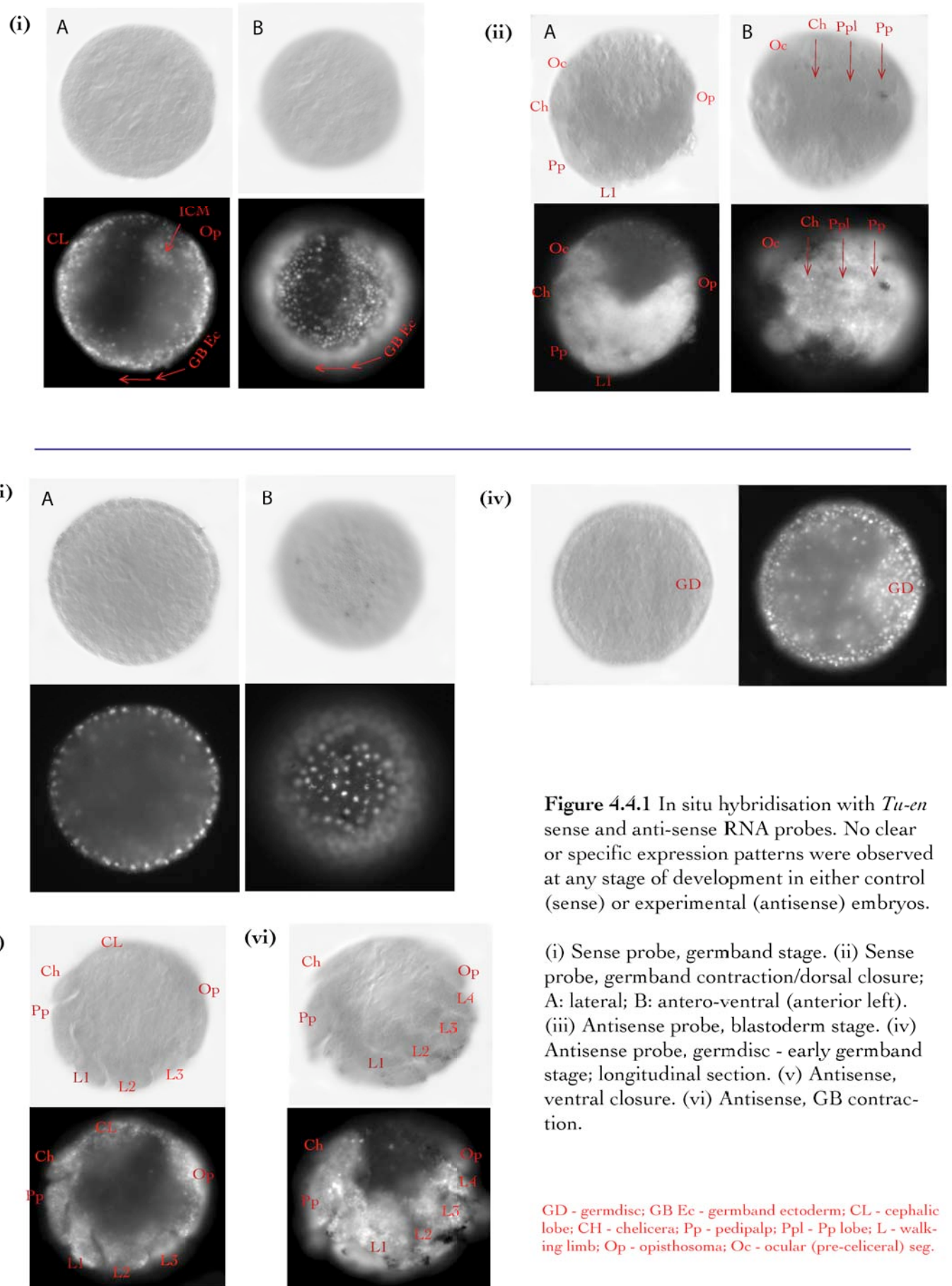
ii) Assessment of failed *Tu-en* mRNA detection

Inability to detect specific staining related to *Tu-en* mRNA transcription may have been due to a number of hypothetical causes:

- degradation of embryonic mRNA during fixation
- degradation of probe mRNA prior to hybridisation
- failure of probe to penetrate embryo cells
- insufficient probe binding due to sub-optimal hybridisation temperature
- insufficient probe-target binding due to sub-optimal probe size
- lack of transcription at developmental stages tested.

RNA degradation

As single stranded RNA is unstable and susceptible to environmental RNases, degradation of target mRNA during fixation or of probe RNA between synthesis and hybridisation is a plausible explanation for negative *in situ* results. However, I reduced potentially damaging steps in the fixation protocol as much as possible (i.e. minimal boiling to detach membranes), and probes were handled with typical anti-RNase precautions. Furthermore, I obtained positive *in situ* hybridisation results for other target genes (e.g. *Dll*, *EGFR*, *Ubx*) in embryos that had been fixed in an identical way, using probes that had been synthesised and handled under the same conditions as the *Tu-en1* and *Tu-en2* probes. This indicates that fixation does not tend to disrupt mRNA structure and that probes are generally well preserved under the anti-RNase conditions I adopted, although selective or partial



degradation cannot be ruled out for any given mRNA sample given the molecular instability and sensitivity of RNA.

Probe penetrance and hybridisation temperature

It is unlikely that the labelled mRNA probes (350bp - 985bp) failed to penetrate the embryos, as they were of comparative size or smaller than probes that had effectively detected transcription of other target genes. (Hydrolysis, when tested previously, had had no noticeable effect, presumably as embryos are very small - 73µm average diameter - and membrane barriers completely removed so there can be no physical benefit to probe fragmentation.) I tested *Tu-en* hybridisation temperatures from 56°C to 60°C, without any observed effect. By contrast, when hybridising against several other genes, a good degree of signal was obtained at all temperatures tested, and 58°C provided the best signal:noise ratio (c.f. *EGFR* and *Ubx in situ* results). It seems unlikely therefore, that hybridisation temperatures were inappropriate for *Tu-en1/en2* probe binding, preventing detection of transcription.

Probe size

In instances where short *Tu-en1* and *Tu-en2* probes (~350bp) were used, it is conceivable that probe either failed to bind adequately to its target, or may have bound but not effectively reported the presence of target mRNA due to low concentration of DIG-labelled nucleotides. However, signal detection was allowed to proceed at low temperatures overnight or for several hours at room temperature, such that one might expect even very low signal to be detectable. Stronger background staining may have obscured any positive signal induced by prolonged exposure, although any such hypothetical staining mediated by the short probes would be questionable given the failure of longer *Tu-en1* probes to report any signs of transcription.

Absence of Tu-en gene activity

Engrailed factors are widely multi-functional, with En/Hh-mediated signaling implicated as important in segmentation and neurogenesis throughout arthropods as well as in more distant, protostome and deuterostome phyla (c.f. Appendix 5 for details)(Abzhanov and Kaufman 2000b; Davis and Patel 2002; Deutsch 2004; Gibert et al. 2000; Hemmati-Brivanlou et al. 1991; Kettle et al. 2003; Patel et al. 1989; Rida et al. 2004; Tautz 2004; Wedeen et al. 1997; Wedeen and Weisblat 1991). This being so, the complete absence of *Tetranychus urticae en-1* and *en-2 in situ* hybridisation signals in all embryos, from late cleavage to late organogenesis, is very unlikely to be due to genuine absence of transcriptional activity. In addition, an independent research group have reported successful detection of *engrailed* transcription in late stage embryos collected from a Canadian population of *Tetranychus urticae*: by kind permission the following section provides a brief review of their findings.

4.4.2 Successful detection of *Tu-en* mRNA transcripts by others

i) Segmental *Tu-en* activity

A single *Tetranychus engrailed* gene was cloned by Dr K. Lee at the University of Western Ontario, and whole mount *in situ* hybridisation on embryos of mixed ages carried out using a corresponding *Tu-en* mRNA probe. Gene expression data were reported for germband contraction/dorsal closure stage embryos (45-50hrs AEL), and I present a sample of this data in Figure 4.4.2 - images gratefully received from Dr K.-Z. Lee. *engrailed* expression appears typical of other arthropods at a similar stage, in that it marks posterior appendage territories, but the transcription pattern is somewhat unclear so it is not possible to say whether anterior or posterior domain borders are sharp or diffuse; a feature that has been used to argue dependence on paracrine (e.g. Wnt) regulation in other species, e.g. *Drosophila*, *Oncopeltus fasciatus*, *Cupiennius salei* (Damen 2002; Ingham and Martinez Arias 1992; Rogers and Kaufman 1996). Lack of gene expression data from earlier embryos leaves the nature of *engrailed* expression during segment formation and limb specification unknown, and any inferences about putative regulation from adjacent Wnt-secreting cells unfeasible.

ii) *Tu-en* activity in non-segmental contexts

Nothing can be proposed regarding possible non-segment polarity functions of the *Tu-en* paralogue prior to germband contraction, but a few observations for the limited data available may suggest roles in anterior appendage morphogenesis and posterior terminal/proctodeum development after germband contraction has commenced. *En* expression is detected in the region of the reduced opisthosoma and proctodeum (Figure 4.4.2, i & iii), and also in patches along the P-D axis of cheliceral and pedipalp lobe appendages (Figure 4.4.2, ii & iv). Expression in chelicerae and pedipalps is stronger where medial fusion occurs, potentially related to directing morphogenesis that generates the stylophore and infracapitulum of the specialised spider mite gnathosoma (Brusca and Brusca 2003; Crooker et al. 2003).

Tentative *Tu-en* expression in terminal structures and during gnathosomal morphogenesis may be connected to conserved functions described further in **Appendix 5** and discussed in Chapter VII. *En* signalling may have had a role in patterning the terminal digestive tract and terminal structures in the common arthropod ancestor, and in diverse morphogenetic events involving induction of differential cell adhesion and changes in cell form (Chipman et al. 2004a; Damen 2002; Davis and Patel 2002;

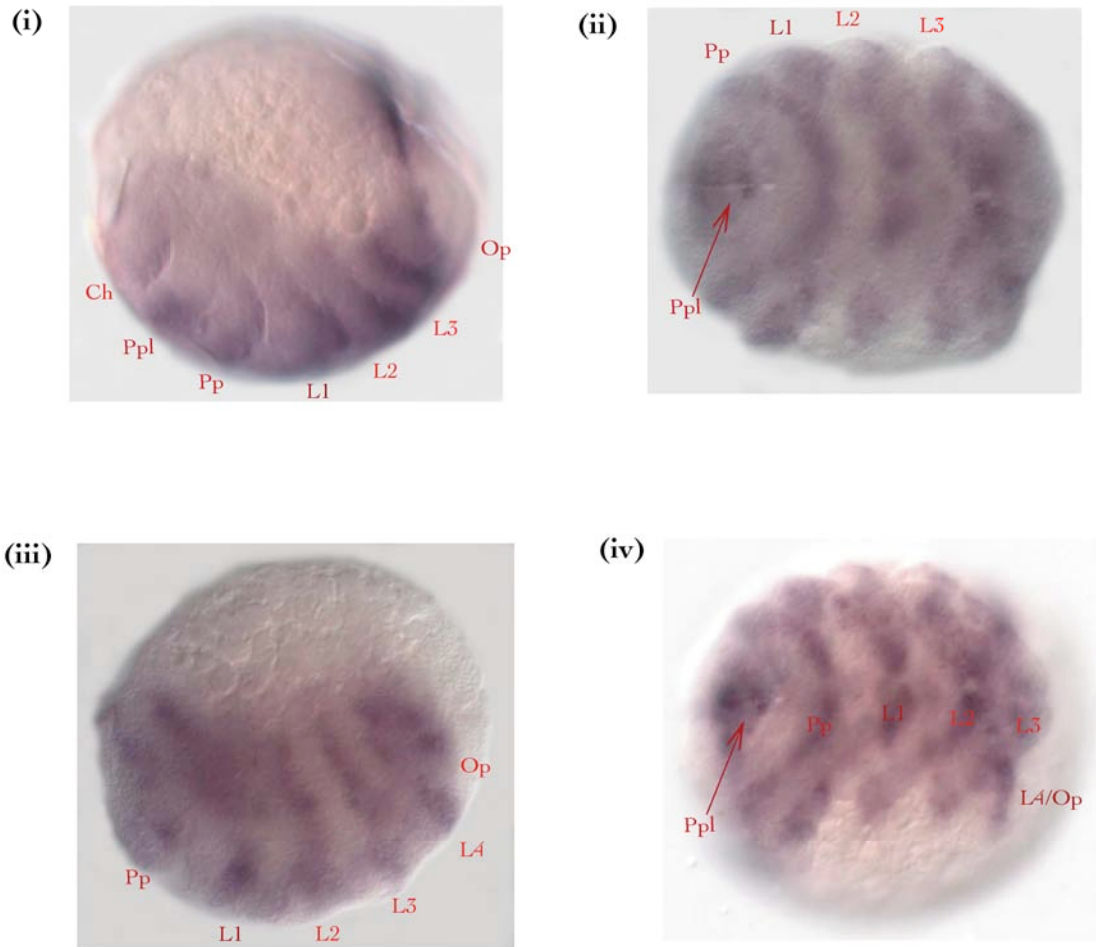


Figure 4.4.2 *Tetranychus engrailed* *in situ* hybridisation signal detected in embryos from late limb development/germband contraction stage (45hrs+). *Tu-en* expression is visible in the posterior part of each segmental appendage, and within the reduced opisthosoma. (i)/(iii): lateral, (ii)/(iv): ventral view. Embryos (i) and (ii) 2-3hrs before (iii) and (iv).

IMAGES FOR THIS FIGURE KINDLY DONATED BY:
Dr Kwang-Zin Lee, Univeristy of Western Ontario

Larsen et al. 2003; Peterson et al. 1998; Simonnet et al. 2004; Wedeen et al. 1997; Wedeen and Weisblat 1991).

iii) Conservation or divergence between *Tu-en* paralogs?

Sequence data for the salient UWO *Tu-en* gene was not available to me, so it is impossible to say which of the two paralogues I cloned, *Tu-en1* or *Tu-en2*, it matches, and therefore which of the two paralogues exhibits posterior segmental appendage expression. It would be ideal to characterise mRNA transcription patterns for both *Tu-en1* and *Tu-en2*, as diverse forms and degrees of functional divergence between *en* paralogs have been documented in several insects, crustaceans and the spider *Cupiennius*, based on comparative gene or protein expression patterns (Peterson, 1998; Dhawan, 2003; Peel, 2006; Abzhanov, 2000; Quéinnec, 1999; Gibert, 2000; Damen, 2002). This issue is discussed further in Chapter VII.

4.5 *Tetranychus* Engrailed protein expression

4.5.1 Monoclonal Engrailed and Invected antibodies 4F11 and 4D9

i) MAb 4F11 and 4D9 synthesis

Engrailed protein expression patterns during *Drosophila* development were described in detail by DiNardo et al (1985) and Karr et al (1989) using rabbit polyclonal and mouse monoclonal antibodies against *Drosophila* Engrailed. These early descriptions detail proposed roles for *engrailed* in control of syncytial nuclear division at early cleavage, and after cellularisation in segmentation, neurogenesis and various types of morphogenesis – from segmental furrow formation to complex head morphogenesis (DiNardo et al. 1985; Karr et al. 1989). Patel et al. (1989) synthesised two monoclonal antibodies (4D9/ α -inv and 4F11/ α -en) which both recognised Engrailed and Invected in *Drosophila*, 4D9 able to recognise Engrailed orthologs in a wide range of metazoans, including insects, crustaceans, molluscs and chordates (Patel et al. 1989; Tautz 2004; Wanninger and Haszprunar 2001). It was demonstrated, *via* homeodomain alignments and targeted deletion constructs assayed

against 4D9, that the 4D9 epitope includes residues 38-48 of the homeodomain and effective binding depends pivotally on glycine (G) at residue 40 (Patel et al. 1989; Wedeen et al. 1997). Various arthropods further to those examined in Patel et al. (1989) have been examined for Engrailed protein expression using MAb 4D9, with successful cross-reaction consistently linked to a conserved glycine at homeodomain residue 40, as summarised in Figure 4.5.1 (Fleig, 1990; Whittington, 1991; Grbic, 1996; Rogers, 1996; Wedeen, 1997; Quéinnec, 1999; Abzhanov, 2000; Dhawan, 2003).

ii) Taxonomic breadth of MAb 4D9 and 4F11 cross-reactivity

Positive cross-reactivity with MAb 4D9 (anti-Inv) has been observed in species including insects *Acheta domestica* (Orthoptera), *Oncopeltus fasciatus* (Hemiptera), *Apis mellifera* (Hymenoptera), *Bombyx mori* (Lepidoptera), *Ctenophalides felis* (Siphonaptera), malacostracan crustaceans *Porcellio scaber* and *Procambarus clarkii*, scaphopod mollusc *Antalis entalis*, and in the chordate *Xenopus laevis* (Dhawan, 2003; Peterson, 1998; Quéinnec, 1999; Rogers, 1996; Abzhanov, 2000; Wedeen, 1997; Fleig, 1990; Wanninger, 2001; Hemmati-Brivanlou, 1991). Monoclonal antibody 4F11 (anti-En) has been less widely tested outside *Drosophila*, and its specific epitope remains unknown (Patel et al. 1989). *Bombyx mori* (Lepidoptera) exhibits *Drosophila*-like engrailed expression, and 4F11 in both *Porcellio* and *Procambarus* detects protein in CNS and PNS cells, primarily in the pereon (Abzhanov and Kaufman 2000b; Dhawan and Gopinathan 2003).

iii) Lack of MAb 4D9 and 4F11 cross-reactivity in *Tetranychus*

Within the EH4/homeodomain of both *Tetranychus* En1 and En2, residue 40 is represented by lysine (K), ruling out cross-reactivity with MAb 4D9 (see Figure 4.5.1 for epitope alignment) (Patel et al. 1989). Indeed, antibody staining of whole embryos with MAb 4D9 or 4F11 did not detect protein expression at any stage, in spite of testing a range of reagent concentrations (1:3 to 1:100) and allowing peroxidase or phosphatase chromogenesis to proceed for a prolonged period (up to 2 hours). Negative results with 4D9 and 4F11 suggest a lack of cross-reactivity, explained by divergence within target Invected/Engrailed epitopes relative to *Drosophila*.

4.5.2 Polyclonal Engrailed antibody 'α-ht-en'

A rabbit polyclonal antibody raised against the Engrailed homeodomain and 27 EH5 domain residues of the leech *Helobdella triserialis* (α-ht-en) also recognises Engrailed in the onychophoran *Akanthokara kaputensis*, and has been used to detect En protein expression patterns in both species (Wedeen et al. 1997; Wedeen and Weisblat 1991). *Helobdella* En has asparagine (N) at homeodomain residue 40, and

	20	40	60	4D9	
DmEn	LKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKSTG	✓			
DmInv	LKHEFNENRYLTEKRRQQLSGELGLNEAQIKIWFQNKRAKIKKSSG	✓			
BmEn	LKHEFAENRYLTERRRQSLAAELGLAEAQIKIWFQNKRAKIKKASG	✓			INSECTA
BmInv	LKHEFAENRYLTERRRQSLAAELGLAEAQIKIWFQNKRAKIKKASG	✓			
AmEn	LKREFAENRYLTERRRQQLSRDLGLNEAQIKIWFQNKRAKIKKASG	✓			
AfEn	LKHEFNENRYLTERRRQDLARELGLHENQIKIWFQNNRAKIKKSSG	✓			
ScEn1	LASEFTDNRYLSEERRQRLARQLGLNESQIKIWFQNKRAKIKKTIP	✓			
ScEn2	LKMEFQONRYLTEKRRQDLAGEQLGLNESQIKIWFQNKRAKIKKTTG	✗			CRUSTACEA
PcEn1	LKKEFQENRYLTEKRRQDLARDLGLNESQIKIWFQNKRAKIKKQAG	✓			
PcEn2	LKKEFEDNKYLTEKRRQDLARELGLNESQIKIWFQNKRAKIKKASG	✓			
SmEn	LKKEFQENRYLTEKRRQDLARELKLNESQIKIWFQNKRAKIKKASG	✗			
LaEn	LKKEFQENRYLTEKRRQDLARELKLNES-----	✗			MYRIAPODA
GmEn	LKKEFQENRYLTEKRRQDLARELKLNESQIKIWFQNKRAKIKKASG	✗			
CsEn1	LKHEFQENRYLTERRRQDLAKDLQGLNESQIKIWFQNNRAKIKKASG	✗			
CsEn2	LRNEFQORSEYLSEAGRRELAQELQLTEAQVKIWFQNNRAKIKKNCG	✗			
AtEn	LRQEFSENRYLTERRRQDMARDLKLNESQIKIWFQNNRAKIKKINP	✗			CHELICERATA
AlEn	LKQEFQENRYLTEKRRQDLAKDLKLNESQIKIWFQNKRAKIKKST-	✗			
TuEn1	LKQEFQDNRYLTEKRRQDLARDLKLNESQIKIWFQNKRAKIKKAGG	✗			
TuEn2	LKQEFTEENRYLTNEKRRQDLARELKLNESQIKIWFQNKRAKIKKTTTS	✗			
AkEn	LKKEFQENRYLTEKRRQDLANDLKLNESQ	✗			ONYCHOPHORA
Csp.En	LKKEFEQENRYLTEARRQELAAELNLNESQIKIWFQNKRAKIKKATG	✗			
PdEn	LKKEFEENRYLAEQRRQELARDLKLNESQIKIWFQNKRAKAKKASG	✗			LOPHOTROCH'A
PvEn	LKHEFDENRYLTETRRQHLANELGLHESQVKIWFQNKRAKIKKATG	✓			
HtEn	LKREFSENKYLTEQRRTCCLAKELNLNESQIKIWFQNKRAKMKKASG	✗			
HeEn	LKQEFQQSNYLTEQRRRTLAKELTLSESQIKIWFQNKRAKIKKASG	✗			
SkEn	LKREFDDSRYLTEQRRQSLAKELKLNESQIKIWFQNKRAKIKKSTG	✗			
OdEn	LQEEFDKHQYLTEDRRVLLAKQLKLSVSQIKVWFQNKRAKIKKVSG	✗			DEUTEROSTOM'A.
CsavEn	LQQEFQKSQYLTEDRRIRVSESLGLSVSQIKVWFQNKRAKIKKTSG	✓			
BfEn	LKKEFQENRYLTEQRRQDLARELKLNESQIKIWFQNKRAKIKKAAG	✗			
DrEn1	LKAEFQTSRYITEQRRQALARELGLNESQIKIWFQNKRAKIKKSSG	✓			
MmEn2	LKAEFQTNRYLTEQRRQSLAQELSLNESQIKIWFQNKRAKIKKATG	✗			
EH4/Homeodomain					

Figure 4.5.1 Multiple alignment of 46 amino acid stretch of the Engrailed homeodomain (EH domain 4) for various protostomes and deuterostomes (residues 16 to 63). The key epitope for positive recognition of *engrailed* genes by monoclonal antibody 4D9 (Patel et al, 1989) is highlighted in **turquoise** (residues 38-48); Residue 40 (in **red**), must be glycine (G) or asparagine (N) for successful epitope binding. Demonstrated or predicted positive (✓) and negative (✗) cross-reactivity with α -En 4D9 is marked for each species on this basis. *Tetranynchus urticae* En1 and En2 should not be detected with the 4D9 antibody, having lysine (K) at 40, which would obstruct binding.

Abbreviations: Cs - *Cupiennius salei*; At - *Achaearanea tepidariorum*; Al - *Archegosetes longesitosus*; Tu - *Tetranynchus urticae*; Sm - *Strigamia maritima*; La - *Litobius atkinsoni*; Gm - *Glomeris marginata*; Af - *Artemia franciscana*; Sc - *Sacculina carcini*; Pc - *Procambarus clarkii*; Bm - *Bombyx mori*; Dm - *Drosophila melanogaster*; Ak - *Akathokara kaputensis*; Csp. - *Chaetopterus sp.*; Pd - *Platynereis dumerilii*; Pv - *Patella vulgata*; Ht - *Helobdella triserialis*; He - *Helicodaris erythrogramma*; Csav - *Ciona savignyi*; Od - *Oikopleura dioica*; Sk - *Saccoglossus kowalevski*; Bf - *Branchiostoma floridae*; Dr - *Danio rerio*; Mm - *Mus musculus*.

Akanthokara lysine (K), explaining why they do not cross-react with MAb 4D9, and why it was necessary to synthesise a species-specific antibody in leech; one that fortuitously detects the onychophoran En ortholog too (see Figure 4.4c). The α -ht-en polyclonal antibody was not available to test in *Tetranychus urticae* but it would be interesting to do so, especially given a degree of sequence similarity in the EH4 domains of *Tetranychus* and *Akanthokara* (e.g. lysine at HD residue 40: Figure 4.4.3), a member of the Euarthropod outgroup Onychophora (Blaxter 2001; Garey et al. 1996; Giribet et al. 1996; Giribet et al. 1999; Simonnet 2005).

Conclusions to Chapter IV

i) *Tetranychus Wg/Wnt1*

Section 4.1 of this chapter has outlined many difficulties encountered in amplifying a *Tetranychus wg/Wnt1* ortholog from genomic DNA, cDNA and RNA templates, in spite of testing many primer combinations, PCR programs and conditions, and both standard and Titanium (hyper-active) Taq polymerases. Three EST sequences with affinity to *Wnt* genes were made available to me and termed *Tu-WntA*, *Tu-WntB* and *Tu-WntC*, but their specific identities within the diverse *Wnt* super-family could not be resolved by Bayesian inference, save to exclude them with high probability from *Wnt5* or *Wnt10* orthology. I conclude that either a *Wnt1* gene remains to be cloned from *Tetranychus urticae*, or that this particular *Wnt* family member has been lost in the spider mite lineage, and its functions partly or fully co-opted by one or more other *Wnt* paralogs, or genes unrelated to *Wnt*. Interestingly, consensus trees resulting from Bayesian analysis of *Tu-Wnts* cast doubt on the assertion based on comparison of *Drosophila*, *Anopheles* and *Caenorhabditis elegans* genomes that representative *Wnt2*, *Wnt5*, *Wnt4*, *Wnt8* and *Wnt11* genes have been lost in the Ecdysozoa, thus pointing out the need for a diverse range of taxa when attempting to examine the evolution of large gene families, as gene loss in a few disparate species may not indicate its absence in their common ancestor.

In the absence of *Wg/Wnt* clones with which to probe spider mite embryos for mRNA transcripts, section 4.2 documents my attempt to uncover domains of protein activity instead, testing for cross-reactivity with an antibody against the Wg morphogen as well as three antibodies against the Wnt signal transduction factor Armadillo/ β -catenin. The antibodies were monoclonal, raised against fragments of *Drosophila* Wg, *Drosophila* Arm or human β -catenin, and although cross-reactive in closely related species (e.g. *Drosophila* Wg and Arm antibodies work in lepidopterans *Bombyx mori* and

Manduca sexta respectively), the only patterns observed in *Tetranychus* embryos were attributed to background staining.

Given the diversity of *Wnt* super-family genes in ecdysozoans, and multiple roles for intercellular Wnt signalling during different phases of arthropod development, I predict that multiple *Wnt* orthologs remain to be characterised in *Tetranychus*, functioning in a range of segmental and non-segmental contexts, such as in early embryo asymmetry, parasegment boundary formation, proximo-distal limb patterning and neural morphogenesis. If the full range of spider mite *Wnt* paralogs can be sequenced and their gene expression patterns described, it would be possible to confirm or refute the expectation of parasegmental *Wnt* gene activity (mediated by a *Wnt1* ortholog, or combination of orthologs whose sub-functionalisation mimics typical arthropod *Wnt1* deployment), and hence assess the role of parasegmental *Wg/Wnt* expression in activating Dll during specification of chelicerate prosomal limbs.

ii) *Tetranychus En* genes

In section 4.3. of this chapter we have seen that the *Tetranychus* genome contains at least two paralogous *engrailed* genes, provisionally termed *Tu-en1* and *Tu-en2*. Preliminary sequences were extended by inverse PCR, allowing confirmation of *Tu-en1* and *Tu-en2* orthology to arthropod *engrailed* genes by Bayesian inference. Alignment of arthropod En protein sequences across diagnostic Engrailed Homology (EH) domains 1 - 5 reveals an 'RX' motif (either RS or RR) at the C' end of the EH2 Exd-binding domain, directly following a conserved SSG motif. An RX motif is absent in all chelicerates so far studied, including *Tetranychus*, but is present in one paralog in some myriapods, some crustaceans and all hexapods. A parsimonious explanation for this is that an ancestral *en* gene duplicated before divergence of major arthropod lineages, and that an RX motif was inserted into one paralog (only) at the base of the Mandibulata, followed by subsequent loss in some myriapod and crustacean lineages but fixation of an RS motif in the Hexapoda: the EH2 RS/RR motif may add an extra Ca²⁺-dependent kinase phosphorylation site, potentially affecting En protein activity.

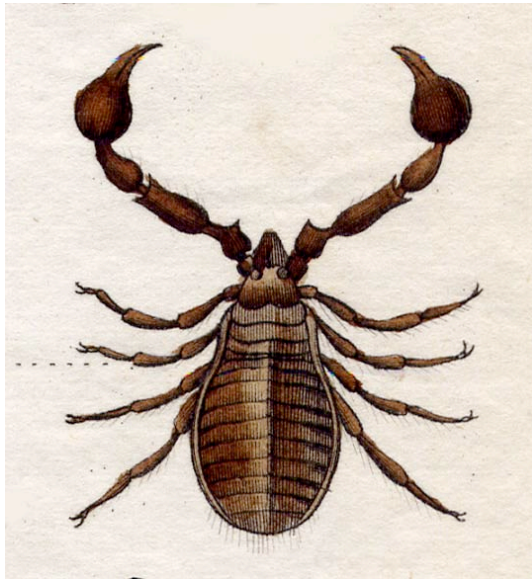
As detailed in section 4.4, I did not manage to positively detect either *Tu-en* mRNA or Tu-En protein in whole mount spider mite embryos. *Tu-en1* and *Tu-en2* ssRNA probes between 350bp and 985bp in length were used for *in situ* mRNA hybridisation, but no gene expression was detected; probe or spider mite RNA degradation are assumed responsible for this failure. Another research group has reported *Tu-en* mRNA expression in late embryos, localised to posterior portions of segmental appendages: this suggests conservation of the segment polarity role of at least one *Tu-en* gene, but whether *Tu-en* genes act earlier to organise para-segmentation or activate *Dll* in limb primordia along the A-P axis, is unknown. I tested for En protein in *Tetranychus* embryos with monoclonal antibodies

against *Drosophila* En paralogs, but antibody staining did not work, probably due to epitope sequence divergence. A polyclonal antibody (α -ht-en) against the Engrailed EH4 + N' EH5 domain of the leech *Helobdella triserialis* cross-reacts in the onychophoran *Akanthokara kaputensis*, whose EH4 domain notably shares many sequence similarities with *Tetranychus* En proteins. Therefore, as well as working to further extend *Tu-en1* and *Tu-en2* gene sequence data and visualise *Tu-en* mRNA transcripts *in situ*, it would be ideal to obtain and test the α -ht-en antibody in spider mite embryos, in the hope of cross-reactivity and detection of *Tu-En* protein throughout development. Specific EH4/EH5 residues critical to α -ht-en epitope recognition are unknown, so any protein domains detected could derive from the activity of either one or both *Tu-En* paralogs. Species-specific antibodies against *Tu-En1* and *Tu-En2*, in addition to mRNA transcription patterns for both genes, are required to distinguish sub- or para-functionalisation between paralogs. As comparative studies indicate ancestral arthropod *engrailed* functions in organising axial segmentation and neurogenesis, as well as in development of terminal structures (e.g terminal digestive tract, caudal circae) and diverse forms of morphogenesis, I would predict seeing *Tu-en1* and/or *Tu-en2* mRNA/protein activity linked with one or more of these processes - if experimental problems with *Tetranychus* can be overcome.

CHAPTER V

REGULATION OF APPENDAGE SPECIFICATION BY CANDIDATE DORSO-VENTRAL PATTERNING GENES:

TETRANYCHUS DPP AND *EGFR*



Faux scorpion: Pince fasciée

Introduction to Chapter V

Why study *decapentaplegic* and *epidermal growth factor receptor*?

As described in Chapter I, between 5-5.5hr AEL in *Drosophila* embryogenesis, dorsal Decapentaplegic (Dpp) and ventral Epidermal Growth Factor Receptor (EGFR) signalling gradients antagonise one another along the D-V axis, mutually repressing each other's transcription. Simultaneously, both Dpp and EGF-mediated signals target the early *Dll* enhancer for repression, restricting *Dll* transcription to ventro-lateral clusters of approximately 20 limb primordium cells, located within the parasegmental competence domain prescribed by En/Wg cues (Gilbert 2000; Lewin 2002). Although Dpp signalling as it relates to dorsal blastoderm, extra-embryonic and proximo-distal limb patterning may be conserved among the arthropods, it seems that outside *Drosophila* there may be no strictly conserved requirement for *dpp* to restrict *Dll* transcription during limb allocation. Patterns of MAPK signalling induced by EGF-R activity have not been described or analysed in any arthropods other than *Drosophila*. In this chapter I characterise spider mite *dpp* and *EGFR* orthologs, aiming to expand the taxonomic range of species in which candidate dorso-ventral *Drosophila Dll* regulatory network orthologs have been examined, and to clarify what conserved roles, if any, *dpp* and *EGFR* orthologs may have played in ancestral chelicerate and arthropod limb specification.

Tetranychus decapentaplegic

Section 5.1 of this chapter presents a spider mite *dpp* ortholog (*Tu-dpp*) that I obtained by degenerate PCR, and extended by inverse PCR and cDNA library screening. Approximately 1.1Kb of the *Tu-dpp* gene was cloned, including 5' untranslated region (UTR), complete coding sequence (CDS), 3' UTR and in cDNA clones, terminal ribosome drop-off site and poly-adenylated tail. *Tu-dpp* was shown by multiple sequence alignment and Bayesian inference to be a true *dpp*/BMP2/4 ortholog, and codes for a protein comprising N' signal sequence, pro-domain with C' basic cleavage residues, and mature ligand domain with 7 conserved TGF- β family cysteine residues. Several ssRNA probes 660bp to 1072bp in size were synthesised from *Tu-dpp* inserts in both pGEM®-T Easy and pBK-CMV vectors, but detection of mRNA transcripts in whole mount spider mite embryos was unsuccessful (section 5.2). An antibody is available against phospho-Smad-1, a signal transducer in the Dpp pathway, and reveals patterns of Dpp signalling activation in a number of species: as described in section 5.3 I tested it in *Tetranychus* embryos and also in the spider *Cupiennius salei*, whose embryos were available at the time and for which *Cs-dpp* mRNA expression patterns are available for

comparison (Prpic et al. 2003). Smad-1 phosphorylation was not convincingly detected in *Tetranychus* embryos, except in two patches at the sites of independent proctodeum and stomodeum invagination, suggestive of an artefact arising from physical properties of tissue layers at that time in morphogenesis. By contrast, α -p-Smad-1 detected Dpp signal transduction well in *Cupiennius*, mimicking *Cs-dpp* mRNA expression, particularly in association with early segmentation, appendage formation and proximo-distal limb development.

Tetranychus Epidermal Growth Factor Receptor

Section 5.4 presents the results of seeking to clone a *Tetranychus Epidermal Growth Factor Receptor* (*EGFR*) ortholog by degenerate PCR, cDNA library screening and inverse PCR. Degenerate PCR yielded 300-350bp fragments of what appeared to be two copies of *Tetranychus EGFR*, and hence were named *TER-a* and *TER-b*. Gene extension of *TER-a* provided approx. 1.5Kb sequence data pertaining to the intracellular tyrosine kinase domain typical of receptor tyrosine kinases (RTKs) such as EGFR. Bayesian analysis showed *TER-a* to be a true spider mite EGFR ortholog, and yet structural comparison indicates that CDS for the N' signal peptide, extracellular domain, transmembrane domain and 5' part of the TK domain remain to be sequenced. Further sequence for the *TER-b* gene could not be obtained by any means, and after repeating original experiments it seems most likely that it derives from another arthropod, possibly a predatory mite whose DNA could have contaminated the DNA preparation from which *TER-a* and '*TER-b*' clones were originally obtained. *TER-a* mRNA transcription was successfully detected *in situ* using ssRNA probes ~680bp to 1Kb in size. Spider mite EGFR transcripts were dynamically expressed during embryogenesis, particularly in association with apparent roles in early ventral fate, ventral neurogenesis, later cephalic neurogenesis and although not evident during early limb specification or development, EGFR is deployed transiently in a medial-proximal telopod segmental domain later in limb development. I tested an antibody against dual-phosphorylated MAPK (α -dpMAPK) in spider mite embryos, aiming to reveal EGFR mediated signalling as dpMAPK is at the end of a ser/thr kinase cascade triggered by EGFR activation, but unfortunately α -dpMAPK does not cross-react in *Tetranychus*.

5.1 Cloning *Tetranychus decapentaplegic*

5.1.1 Degenerate PCR screening for *∂pp*

PCR amplification conditions

I amplified 165bp of a putative *Tetranychus ∂pp* ortholog (*Tu-∂pp*) from genomic DNA, using a touchdown PCR program (see Chapter VIII: Materials & Methods, section 8.2). Gradual decline in temperature during the first 6 PCR cycles creates highly stringent early binding specificity, increasing the likelihood of degenerate primer binding to the desired target DNA sequence, followed by lower temperature amplification of products obtained from the initial stringent selection phase. I tested 10 different degenerate primer pairs; degdpp-F2 (5'-GGC GGC ACC CCA TGT AYG TNG AYT T-3') and degdpp-R1 (5'-GGG TCT GCA CCA CGG CRT GRT TNG T-3') successfully amplified DNA close to the strict predicted size of 153bp (see Appendix 2.5 for all primer sequences). Degenerate primers were targeted against codons for conserved amino acids flanking 2 of 7 signature TGF-β family cysteine residues (Kingsley 1994; Sanchez-Salazar et al. 1996; Yamamoto et al. 2004).

Tu-∂pp sequence data and analysis

The putative *Tu-∂pp* nucleotide sequence codes for part of a mature, C terminal TGF-β ligand domain (Figure 5.1.1,i-ii), including 2 cysteine residues - part of a diagnostic 'cysteine knot' structure - and a putative N-linked glycosylation site (asn-ala-thr), potentially involved in mediating protein secretion (Alberts et al. 1994; Kingsley 1994; Sha et al. 1989). Multiple alignment of available *Tu-Dpp* amino acids with Dpp or BMP2/4 orthologs from a range of metazoan taxa confirms the conserved cysteine and N-glycosylation sites, and supports the true orthology of the *Tu-∂pp* gene identified (Figure 5.1.1,iii). Gene sequence data was tested for similarity to other genes *via* BLAST searches, indicating closest identity to arthropod *∂pp* + vertebrate *BMP2/4* and secondarily to the *∂pp*-related *Drosophila* genes *gbb* (*glass bottom boat*) and *scw* (*screw*) (Altschul et al. 1990; Arora et al. 1994; Decotto and Ferguson 2001; Ray and Wharton 2001).

5.1.2 Further *Tu-∂pp* data from inverse PCR and cDNA library screening

i) Inverse PCR: CDS extension

With the aim of extending the preliminary *Tu-∂pp* DNA fragment obtained by degenerate PCR, I carried out inverse PCR (iPCR) on genomic DNA circularised after restriction enzyme digestion

(i)

```
1/1 31/11
CGC CGG AGA CCC ATG TAC GTC GAT TTT AGT GAA GTT GGT TGG AAC AAT TGG ATT
arg arg arg pro met tyr val asp phe ser glu val gly trp asn asn trp ile
61/21 91/31
GTA GCT CCA CTT GGT TAT CAA GCA TTT TAT TGT GCG GGT GAC TGT CCT CAT CTC
val ala pro leu gly tyr gln ala phe tyr cys ala gly asp cys pro his leu
121/41 151/51
CTC AAC GAT GTC CAT AAT GCC ACA AAT CAC GCC ATT GTT CAA AAT
leu asn asp val his asn ala thr asn his ala ile val gln asn
```

(ii)

```
Fwd→ TuDpp ←Rev
RRRPMYVDFSEVGWNNWIVAPLG YQAFY CAGDCPHLLNDVHNATNHAIVQN

Fwd-degF2 5'-GGC GGC ACC CCA TGT AYG TNG AYT T-3'
Rev-degR1 5'-GGG TCT GCA CCA CGG CRT GRT TNG T-3'
```

(iii)

DrosophilaDpp	RRHSLYVDFSDVGWDDWIVAPLG YDAYYCHGKCPFPLADHFNSTNHAIVQT	Arthropoda
AnophelesDpp	QRKPLYVDFSDVGWDDWIVAPPGYEAYYQGD CRFPIADHLNTNHAIVQT	
PrecisDpp	QRRPLFVDFADVGWSDWIVAPHGYDAYYQGD CPFLSDHLNGTNHAIVQT	
OncopeltusDpp	--HPLYVDFKDVGWDDWIVAPPGYEAFYCHGDC PFPLADHLNSTNHAIVQT	
TriboliumDpp	RRRQMYVDFGSVGWDDWIVAPLG YDAYYCGGCEYPIPDHMTNHAIVQS	
SchistocercaDpp	RRHPLYVDFREVGWDDWIVAPPGYEAWYCHGDC PFPLSAHMNSTNHAIVQT	
CupienniusDpp	RRHALYVDFYDVGWDDWILAPPGYDAYFCHGDC PFPLPDHLNATNHAIVQT	
AchaearaneaDpp	RRHALYVDFSDVGWDDWIIAPPGYNAYFCHGDC PFPLPDHLNSTNHAIVQT	
TetranynchusDpp	RRRPMYVDFSEVGWNNWIVAPLG YQAFY CAGDCPHLLNDVHNATNHAIVQN	
IlyanassaDpp/BMP	RRHALYVDFQEVGWEDWIVAPDGYNAYFCQGD CNFPLAQHLNSTNHAIVQT	Loph'zoa
PatellaDpp/BMP	KRHVLYVDFGDVGWDDWIVAPPGYNAYFCRGECPFPMGQHLNSTHHAIVQT	
Strongylo' BMP2/4	RRHELYVDFSDVHWDDWIVAPAGYQAYYCRGEC PFPLAEHLNTNHAIVQT	Deuterost'a
LytechinusBMP2/4	RRHPLYVDFSDVHWDDWIVAPAGYQAYYCHGEC PFPLAEHLNTNHAIVQT	
TripeustesBMP2/4	RRHPLYVDFSDVHWDDWIVAPAGYQAYYCHGEC PFPLAEHLNTNHAIVQT	
BranchiostomaBMP2/4	RRHSLYVDFSDVGWDDWIVAPPGYQAYYCHGEC PFPLADHLNSTNHAIVQT	
DanioBMP2/4	RRHALYVDFSDVGWNEWIVAPPGYHAFYCHGEC PFPLPDHLNSTNHAIVQT	
HomoBMP2	RRHSLYVDFSDVGWDDWIVAPPGYQAFYCHGDC PFPLADHLNSTNHAIVQT	
HomoBMP4	RRHSLYVDFSDVGWDDWIVAPPGYQAFYCHGDC PFPLADHLNSTNHAIVQT	Porifera
MusBMP4	RRHSLYVDFSDVGWDDWIVAPPGYQAFYCHGDC PFPLADHLNSTNHAIVQT	
AcroporaDpp	QRHPLYVDFSEVGWDDWIVAPPGYQGFYCKGEC PFPIADHLNTNHAIVQT	

Figure 5.1.1 (i) Nucleotide and amino acid sequence for putative *dpp* obtained by degenerate PCR, representing C terminal piece of secreted TGF- β ligand domain. **(ii)** Tu-Dpp protein overview and degenerate primer sequences. Primer target regions (grey); 34 novel amino acids including 2 conserved TGF- β -family cysteine residues (blue) and possible N-linked glycosylation site (red). **(iii)** Multiple sequence alignment (51 amino acids) for Dpp/BMP2/4 orthologs in a range of taxa. Bold: *Tetranynchus urticae* Dpp; blue – cysteine knot residue; red font – N-linked glycosylation site. Abbreviation: Strongylo' – *Strongylocentrotus purpuratus* (Echinodermata).

with DraII, EcoRV or RsaI (see in Chapter VIII Materials and Methods, section 8.2.3 for protocol and PCR conditions (Averof 1994)). I designed three pairs of outward facing primers, specific to the short *Tu-dpp* sequence already cloned. Primer combination U61/L56 for the first iPCR and U111/L33 for ‘nested’ PCR amplified DNA that was shown to contain new *Tu-dpp* material (for a complete list of iPCR primers sequences, see Appendix 2.5). Genomic *Tu-dpp* approaching 450bp (444bp) in length was obtained in total, adding 279bp to the original 165bp degenerate PCR fragment: new coding sequence was added mostly in a 3’ direction, reaching a TAA termination codon and 156bp beyond (see Figure 5.1.2a,i).

ii) cDNA library screen: full length *Tu-dpp* transcript

Gene sequence and structure

I sequenced 1071bp of *Tu-dpp* mRNA from the transcriptome, following successful phage isolation from an embryonic cDNA library and excision into pBK-CMV phagemid vector (details in Materials & Methods - Chapter VIII). The transcript recovered (Figure 5.1.2a,i-ii) includes 663bp coding sequence, 45bp 5’ UTR and 360bp 3’ UTR, up to and including a poly-adenylation signal (AATAAA) and poly-A tail.

The general structure of TGF- β /Dpp proteins includes an **N’ signal sequence** (15-25aa) whose cleavage initiates export of a **pro-domain** + **ligand domain** precursor, modified and cleaved before **mature ligand** secretion (Kingsley 1994; Sanchez-Salazar et al. 1996; Sha et al. 1989). Given the typical structure of signal sequence domains (N’-basic, central hydrophobic and C’-polar), the putative *Tu-Dpp* N’ export signal sequence (Figure 5.1.2a) may be cleaved at a polar serine (aa #25) that follows 9 hydrophobic residues (Kingsley 1994; von Heijne 1986). However, the *Tu-Dpp* N’ polypeptide does not follow a strict basic-hydrophobic-polar structure, so this signal peptide ‘cleavage site’ remains tentative.

According to Sánchez-Salazar et al. (1996), both insect and vertebrate *Dpp/BMP2/4* genes have a conserved intron splice site within the N’ pro-domain, containing early-acting regulatory elements (Sanchez-Salazar et al. 1996). In spite of implied existence before pro- and deuterostome divergence, I could not locate conserved trans-splice site coding sequence to my satisfaction in *Tu-Dpp*. The best identifiable approximation to an introns splice site alignment (below) is far from ideal, although it is within the putative N’ pro-domain region:

Tu-Dpp	PPFTIVSGHHD<D>SKSTDSN	
Tc-Dpp	ANTIRSFTHV<A>SPIDEKF	
Dm-Dpp	ANTVRSFTHK<D>SKIDDRF	
Hs-BMP2	ANTVRSFHHE<E>SLEELEP	
HS-BMP4	ANTVRSFHHE<E>HLENIPG	Yellow = residues putatively shared with <i>Tu-Dpp</i>

(i)

```
1
CCG GCA CGA GGT TTA CTA ACA CCA ATA TCA TCT TGA TGA GTG ACC ATG ACT TTT CAA TGG
61/6 91/16
ACC CTT GTG GCA ACA ATT TTC TAC ATT GAG ATT GCG GTT TGT CCT CCG TTC ATC GTA TCT
thr leu val ala thr ile phe tyr ile glu ile ala val cys pro pro phe ile val ser
121/26 151/36
GGT CAC CAT GAC GAT TCA AAA TCA ACA GAT TCT AAT TTA TTT AAC CAA TAT CAT CTG AAT
gly his his asp asp ser lys ser thr asp ser asn leu phe asn gln tyr his leu asn
181/46 211/56
TGG ACG AGT CTG GCT CTT GAC GAT CCA AGC CAA TAT TTT TAC ACC ACC AAT TTT AAA ACT
trp thr ser leu ala leu asp asp pro ser gln tyr phe tyr thr thr asn phe lys thr
241/66 271/76
GAG GAG AAT GAA CCT TTG CTA GAA GCA CCT TTA CTA GTG ACT TAT TCA AAT GAT GGC ACA
glu glu asn glu pro leu leu glu ala pro leu leu val thr tyr ser asn asp gly thr
301/86 331/96
CCT GAT GCC TCA AAT GTT TTA AAA CGT GAT AAG AGA CAA GTT GAT CAA AGA AAT AAG CCA
pro asp ala ser asn val leu lys arg asp lys arg gln val asp gln arg asn lys pro
361/106 391/116
GTC AGT GGA CGT AAG AGT CGT CGC AAA CAA AGG AAA GAG GCA TGT CGC CGG AGA CCC ATG
val ser gly arg lys ser arg arg lys gln arg lys glu ala cys arg arg arg pro met
421/126 451/136
TAC GTC GAT TTT AGT GAA GTT GGT TGG AAC AAT TGG ATT GTA GCT CCA CTT GGT TAT CAA
tyr val asp phe ser glu val gly trp asn asn trp ile val ala pro leu gly tyr gln
481/146 511/156
GCA TTT TAT TGT GCG GGT GAC TGT CCT CAT CTC CTC AAC GAT GTC CAT AAT GCC ACA AAT
ala phe tyr cys ala gly asp cys pro his leu leu asn asp val his asn ala thr asn
541/166 571/176
CAC GCC ATT GTT CAA AAT TTA GTT AAC TCT GTT ACC CCT AAC AAG GTT CAA AAG GCT TGC
his ala ile val gln asn leu val asn ser val thr pro asn lys val gln lys ala cys
601/186 631/196
TGT GTT CCC ACC GAG CTT TCA TCA ATT TCA ATT CTT TAT GTC GAC GAA TAT GGA AAA GTT
cys val pro thr glu leu ser ser ile ser ile leu tyr val asp glu tyr gly lys val
661/206 691/216
GTT CTC AAA AAT TAT CAA GAC ATG GTA GTG GAA GCT TGT GGC TGC CGG TAA TCA ACT ATT
val leu lys asn tyr gln asp met val val glu ala cys gly cys arg - -
721 751
GAC AAA ATC ACT GCC AAT TCA CAA CAA AAC ATC AGC CAA AGG TGT AAA ATA TCA ATT GTA
781 811
AAA TTG TAA ACA CAA ATG GAG CCT CTT ATC TAG GCC TTC AAG ATT ATA GTC AAT TGC CAT
841 871
AAA TGC CTT TTG TTT GCC ATT TTG TAC TCG AGT TTG AAC AAC TCA ATT TTA ATC AAT TAG
901 931
CAA ACA AAA ACT ATT TTA CTT TAA TTA TTC TGA TCG CTC TTG AAA TCG AGT ATC AAC CAA
961 991
ACC AAA CAC AAA AGT AAA AAC AAA AGC AAG GTT TGT GTA ATT ATA AAT TGA ATA AAA ACC
1021 1051
AAA AGG TAA ACC AAA TAA ACA CAT AAA CAA AGA AAA AAA AAA AAA AAA AAA
```

(ii)

```
5' MTFQWTLVATIFYIEIAVCPFFI VS GHDDSKSTDSNLFNQYHL NWT SLALDDPSQYFYTTNFKTEENEPLLEAPLLVT
YSNDGTPDASNVLKRDQRQVDQRNKPVS GKSRRK QRKEA RRRPMY VDFSEVGWNNWIVAPLGYQAFY CAGDC PHLLNDV
H NAT NHAIVQNLVNSVTPNKVQK ACCV TELSSISILYVDEYGVVLKNYQDMVVEAC GR 3'
```

Figure 5.1.2a (i) Nucleotide sequence of *Tudpp* CDS, retrieved by screening cDNA (λ -Zap) library, with amino acid translation where appropriate. Bold: *Tudpp* sequence also obtained by inverse PCR (prior to library screening). **(ii)** summary of Dpp protein code, iPCR sequence again in bold. Grey: 5' (U24) and 3' (L111) primers, used to (re-)amplify screening probe DNA from original degenerate PCR clones. \odot : stop codon. Yellow: signal sequence cleavage site. Blue: structural cysteine residues. NXT: N-linked glycosylation site. RKSRRK: basic pro-domain C' cleavage site.

A region of basic residues, functionally arranged as either di-basic, tri-basic or an R/K-X-X-R motif (Figure 5.1.2a), occurs at the presumed site of C' pro-domain removal by subtilisin family proteases (Alberts et al. 1994; Barr 1991; Lepage et al. 1992). Two possible sites of N-mediated glycosylation, implicated in early oligosaccharide addition to precursor protein, are present in *Tu*-Dpp; one in the N' pro-domain and one in the C' ligand region (Alberts et al. 1994; Sha et al. 1989). Conserved TGF- β family cysteine residues are also present (Figure 5.1.2a), cys1 – cys6 forming a 'cysteine knot' and cys7 allowing formation of homo- or hetero-dimers *via* a disulphide bridge to cys7 of another TGF- β monomer (Kingsley 1994; Yamamoto et al. 2004). The cysteine knot confers high resistance to extremes of temperature and pH, and ligand dimerisation is essential to activate signalling at specific trans-membrane receptors (Gilbert 2000; Patruno et al. 2002; Persson et al. 1998; Whitman and Raftery 2005).

Multiple protein sequence alignment and Phylogenetic analysis

A comparison of 128 homologous amino acids from Dpp/BMP2/4 orthologs and more distantly related proteins (Figure 5.1.2b) reveals the seven strikingly conserved cysteine residues and N-linked glycosylation site, that identify *Tu*-Dpp and the other orthologs as TGF- β family members (Alberts et al. 1994; Sanchez-Salazar et al. 1996).

I carried out Bayesian inference analysis of a 384bp nucleotide dataset, corresponding to CDS for the domain aligned in Figure 5.1.2b. Within the consensus topology, *Tetranychus dpp* receives maximal posterior probability support as a *dpp/BMP2/4* ortholog, and *BMP5/8* + *gbb/scw* genes form a well supported (0.99) outgroup to the *dpp+BMP2/4* clade (Figure 5.1.2c). However, the monophyly of the *dpp+BMP2/4* clade is disrupted by long branch *Nodal* sequences, grouped together as a sub-clade closest to Echinoderm, Lower Chordate and *Acropora dpp* orthologs (Figure 5.1.2c). *Nodal* genes may have evolved by duplication from an ancestral *BMP2/4*-like gene in the Euchordate lineage (*Nodal* orthologs are only known in Euchordates to date), but it is not possible to examine this hypothesis as the evident long branches may have disrupted the topology, making phylogenetic placement unreliable (Kingsley 1994; Whitman and Raftery 2005). In order to remove any adverse long branch effects, I repeated the Bayesian analysis without *Nodal* genes; the resultant topology (Figure 5.1.2d) displays little or no change in probability support values at major nodes and *Tu-dpp* branches within a monophyletic, maximally supported *dpp/BMP2/4* clade.

Dpp/BMP2-4 orthologs:

CsDpp	FQWSEHEPMLIIYSRDPNARTKRRDN	QRRHALYVDFYDVGWDDWI	IAPPGYDAYF	CHGDC	PPFLPDLH	NA	TNHAIVQTLVLSANPAAV	PRAC	CCVPEMSAISILHRDQNDMVVLSVYRDMVVEGCGCR		
AtDpp	HRWNOVEPLLVIYSFNFTNRTRKRDN	QRRHALYVDFSDVGWDDWI	IAPPGYNAYF	CHGDC	PPFLPDLH	NT	TNHAIVQTLVLSANPAAV	PRAC	CCVPTELSPISMLYKDFDNVVLKNYQDMVVEGCGCR		
TuDpp	NEFLLEAPLLVIYSNDGTPRNKRKEA	CRRRPMYVDFSEVGWNNWI	VAPLGQAIFY	CAGDC	PHLLNDLNS	TNHAIVQNLVLS	TNPKVKQA	CCVPTELSSISITLYVDEYGVVLKNYQDMVVEACGCR			
GmDpp	-----	-----	-----	-----	-----	-----	-----	-----			
TcDpp	PQWYQHQP	LLFTYTDDGKNPKRLKDP	QRRRQMYVDFSGVGWDDWI	VAP	LGDAYY						
SgDpp	-----	-----	-----	-----	-----	-----	-----	-----			
OfDpp	-----	-----	-----	-----	-----	-----	-----	-----			
AtDpp	ESWTANRP	FLFTYTDDGRNPARGRENCRRHPL	YVDFYDVGWDDWI	VAP	PGYDAFY	CHGDC	PPFLADHL	NS	TNHAIVQTLVSTNPNI	VPKA	
PcDpp	EDWRVQ	PLLMLYTDEERANKAREICRRPL	FVDFADVGSDWI	VAP	HGYDAYY	CGDC	PPFLSDHL	NG	TNHAIVQTLVSNPAAV	PKA	
DmDpp	ERWQHK	QPLLFTYTDDGRHNKHDDT	RRHSLYVDFSDVGWDDWI	VAP	LGDAYY	CHGDC	PPFLADHL	NS	TNHAIVQTLVSNPAAV	PKA	
AgDpp	DWVQKQ	PLLFTYTDDGRH---	ELCQRKPLYVDFSDVGWDDWI	VAP	PGYEAYY	CGDC	PPFLADHL	NT	TNHAIVQTLVSNP	LAPKA	
IoDpp	ERWRMER	PLLVITDDGKGRSKDKNQ	QRRHALYVDFQEVGWDDWI	VAP	PGYNAYF	CGDC	PPFLAQHL	NS	TNHAIVQTLVSDPTAV	SKA	
PvDpp	QWQIQ	IRPPLVITDDGRRRRNKNE	QRRHLYVDFDVGWDDWI	VAP	PGYNAYF	CGEC	PPFLMGHL	NS	THHAWQTLVSDPTAV	PRAC	
SpBMP2/4	DRWFTRP	QIVTYSDDGRTKKRLKAN	RRHLYVDFSDVHWNWI	VAP	PAGYQAYY	CHGEC	PPFLAEHL	NT	TNHAIVQTLVSNPALV	PKA	
LvBMP2/4	ERWFQTR	PQIVTYSDDGRTKKRLKAN	RRHPLVDFSDVHWNWI	VAP	PAGYQAYY	CHGEC	PPFLAEHL	NT	TNHAIVQTLVSNPALV	PKA	
TbBMP2/4	ERWFTTR	PQIVTYSDDGRTKKRLKAN	RRHPLVDFSDVHWNWI	VAP	PAGYQAYY	CHGEC	PPFLAEHL	NT	TNHAIVQTLVSNPALV	PKA	
BfBMP2/4	HSWQHR	PLLVITDDG--	ROKLKAN	RRHSLYVDFSDVGNWDDWI	VAP	PGYQAYY	CHGEC	PPFLADHL	NS	TNHAIVQTLVSNPALV	PKA
DanioGBMP24	DSWAQ	ARPLLVITYSHDQGRGRGQRNS	QRRHALYVDFSDVGWNNWI	VAP	PGYHAFY	CHGEC	PPFLPDLH	NS	TNHAIVQTLVSN--	SNIPKA	
HomoBMP2	GNWAQ	LRPLLVITFGHDGRGKHHKKN	QRRHSLYVDFSDVGWDDWI	VAP	PGYQAFY	CHGDC	PPFLADHL	NS	TNHAIVQTLVSN--	SSI	
HomoBMP4	GNWAQ	LRPLLVITFGHDGRGKHHKKN	QRRHSLYVDFSDVGWDDWI	VAP	PGYQAFY	CHGDC	PPFLADHL	NS	TNHAIVQTLVSN--	SSI	
MusBMP4	GDWAQ	LRPLLVITFGHDGRGKHHKKN	QRRHSLYVDFSDVGWDDWI	VAP	PGYQAFY	CHGDC	PPFLADHL	NS	TNHAIVQTLVSN--	SSI	
AmBMP /Dpp	DGTIDNK	PPLLVTFSHRTERRSEVNT	QRRHPLYVDFSEVGWDDWI	VAP	PGYQGYF	CHGEC	PPFLADHL	NT	TNHAIVQTLVSNPNNV	PAC	

Dpp/BMP2-4 related orthologs:

HomoBMP5	QGPQSKQ	FMVAFFKASEVSSHQKQA	CKKHLYVSFRDLGWDDWI	IAP	EGYAAAFY	DGEC	SPFLNAHMA	TNHAIVQTLVLMF	PDHVPK	CCAPT	KLNAISVL	YFDDSSNVIL	KKYRNMMVRS	CGCH																																																																							
HomoBMP6	DGPYDKQ	FMVAFFKVEVTSQSLKTA	CKKHLYVSFDLGWDDWI	IAP	EGYAAANY	DGEC	SPFLNAHMA	TNHAIVQTLVLMN	PYVPK	CCAPT	KLNAISVL	YFDDSSNVIL	KKYRNMMVRS	ACGCH																																																																							
HomoBMP7	HGPQNKQ	FMVAFFKATEVPMNQROA	CKKHLYVSFRDLGWDDWI	IAP	EGYAAAYY	DGEC	APFLNSYMA	TNHAIVQTLV	INPETVPK	CCAPT	QLNAISVL	YFDDSSNVIL	KKYRNMMVRS	ACGCH																																																																							
HomoBMP8	RAPRSQQ	PVVTFRASPSPSLPGRQV	CKKHLYVSFDLGWDDWI	IAP	QGYSAAY	DGEC	SPFLNSCMA	TNHAIVQTLVLMK	PNAVPK	CCAPT	KLNAISVL	YFDDSSNVIL	KKYRNMMVRS	ACGCH																																																																							
MusBMP5	HGPQSKQ	FMVAFFKASEVSSHQKQA	CKKHLYVSFRDLGWDDWI	IAP	EGYAAAFY	DGEC	SPFLNAHMA	TNHAIVQTLVLMF	PDHVPK	CCAPT	KLNAISVL	YFDDSSNVIL	KKYRNMMVRS	CGCH																																																																							
MusBMP7	HGPQNKQ	FMVAFFKATEVPMNQROA	CKKHLYVSFRDLGWDDWI	IAP	EGYAAAYY	DGEC	APFLNSYMA	TNHAIVQTLV	INPETVPK	CCAPT	QLNAISVL	YFDDSSNVIL	KKYRNMMVRS	ACGCH																																																																							
MusBMP8	QAPRSQ	QFMVTFRASQSLPGRV	CKKHLYVSFRDLGWDDWI	IAP	QGYSAAYY	DGEC	APFLNSCMA	TNHAIVQTLVLMH	SPSHVPK	CCAPT	KLNAISVL	YFDDSSNVIL	KKYRNMMVRS	ACGCH																																																																							
SpBMP5/8	GNNEG	RPFMVVFQORNEERKPSDWQ	CKRNFLVNFEDLDQWMI	IAP	LGYYAFY	DGEC	APFLNGHMA	TNHAIVQTLVLMH	SPSHVPK	CCAPT	KLNAISVL	YFDDSSNVIL	KKYRNMMVRS	ACGCH																																																																							
DmGbb	KYVDE	FQFPMIGFFRGPELKRKSTRS	COMQTLYIDFKDLGWDDWI	IAP	EGYGAFFY	DGEC	APFLNAHMA	TNHAIVQTLVLMH	PEKRVKPC	CCAPT	RLGALPVL	YHLNDENVNL	KKYRNIMVKS	CGCH																																																																							
DvGbb	KYVDE	NOFPMIGFFRGPELKRKSTRS	COMQTLYIDFKDLGWDDWI	IAP	EGYGAFFY	DGEC	APFLNAHMA	TNHAIVQTLVLMH	PEKRVKPC	CCAPT	RLGALPVL	YHLNDENVNL	KKYRNIMVKS	CGCH																																																																							
AgGbbA	RGSEY	QFLVYVANSQQQRKSQHS	CRIQOLYVSFKDLQWHMI	IAP	EGYGAFFY	DGEC	APFLNAHMA	TNHAIVQTLVLMH	PEKRVKPC	CCAPT	RLGALPVL	YHLNDENVNL	KKYRNIMVKS	CGCH																																																																							
DmScw	ASRTSLEP	FI	VGYNGLRAGPPQS	GERL	NFTVDFKELHMHNMW	IAP	KKFEAFY	GGCC	NFLGTMNA	TNHAIVQTLVLMK	QP-HL	PKPCC	CVPTVL	GAITILRYL	NEDI	IDL	TKYQKAVARE	CGCH																																																																			
BfNd1	IKLSE	VS	LVVFSQDQKADRHDE	DT	PKKVEFWDF	PHIGW	TWII	YPKRFNAFR	CGV	CP	TPVD	QLYHPT	SHAVMTS	ILHKPGK-AP	MPCC	I	PTK	KL	ALS	ML	YLEH-GEV	VL	RH	HED	MI	V	DE	CG	Q																																																								
DanioNd1	NRAK	TSL	IRTA	EHKS	YVARR	NKKPL	CKK	VD	MMVDF	QIG	SD	WII	VY	PKRYNAFR	CG	SC	PP	VD	ET	FPT	TN	HAY	Q	SL	L	Y	PHR-VP	CL	SC	VP	TR	L	AP	LS	ML	YYEN-G	K	M	V	M	R	H	H	E	K	D	M	I	V	E	E	C	G	C	L																														
MusNd1	ROLG	GA	T	L	WEA	ES	SWRA	QRR	HRS	Q	J	OR	V	K	F	Q	V	D	F	N	L	I	G	W	S	W	I	I	V	P	K	Q	N	A	Y	R	CG	E	C	N	V	M	M	Q	Q	K	T	S	L	L	Y	F	E	N-G	E	V	Y	Q	R	D	H	E	E	M	V	E	E	C	G	C	L														
HrNd1	NLM	D	T	T	W	E	R	L	N	R	F	R	E	S	G	H	R	R	T	P	V	H	G	R	K	V	A	F	E	D	F	K	I	G	W	E	M	I	I	V	P	R	R	Y	N	A	F	R	CG	E	C	N	V	M	M	Q	Q	K	T	S	L	L	Y	F	E	N-G	E	V	Y	Q	R	D	H	E	E	M	V	E	E	C	G	C	L		
LvNd1	SPL	T	V	N	D	V	T	L	V	F	S	R	A	P	S	R	R	S	R	S	G	P	Q	R	R	V	D	M	D	V	D	F	G	R	I	G	W	E	M	I	I	V	P	K	Q	F	N	A	Y	R	CG	E	C	N	V	M	M	Q	Q	K	T	S	L	L	Y	F	E	N-G	E	V	Y	Q	R	D	H	E	E	M	V	E	E	C	G	C	L

Figure 5.1.2b Multiple alignment of 127 homologous amino acids from 22 Dpp+BMP2/4 orthologs and 17 Dpp-related orthologs from a range of protostomes, deuterostomes and a poriferan. The residues compared reflect characters included in Bayesian analysis of *Tu*dpp CDS; Dpp-related sequences are included to serve as outgroups. **Blue**: conserved TGF-β family cysteine residues. **Red**: N-linked glycosylation motif. **Green**: basic residue domain, potential pro-domain cleavage region. **Bold**: *Tetranylebus* Dpp. [Abbreviations: At=*Achaezarina tepidariorum*; Cs=*Cupimnia salis*; Tu=*Tetranylebus urticae*; Gm=*Glomeris marginata*; Tc=*Tricholium castaneum*; Sg=*Schistocerca gregaria*; Of=*Oncopeltus fasciatus*; Ar=*Albakea ruaz*; Pc=*Precia coeniza*; Dm=*Drosophila melanogaster*; Dv=*Drosophila gambiæ* (Arthropoda); Io=*Ilyanassa obsoleta* (Mollusca); Sp=*Strongylocentrotus purpuratus*; Lv=*Lytechinus variegatus* (Echinodermata); Tg=*Tricnutes gracilis*; Bf=*Branchiostoma floridae*; Hr=*Halocynthia roretzi* (Chordata); Am=*Acropora millepora* (Porifera); Gbb – Glass bottom boat; Scw – Screw; Nd1 – Nodal; Dpp – Decapentaplegic; BMP – Bone Morphogenic Protein.]

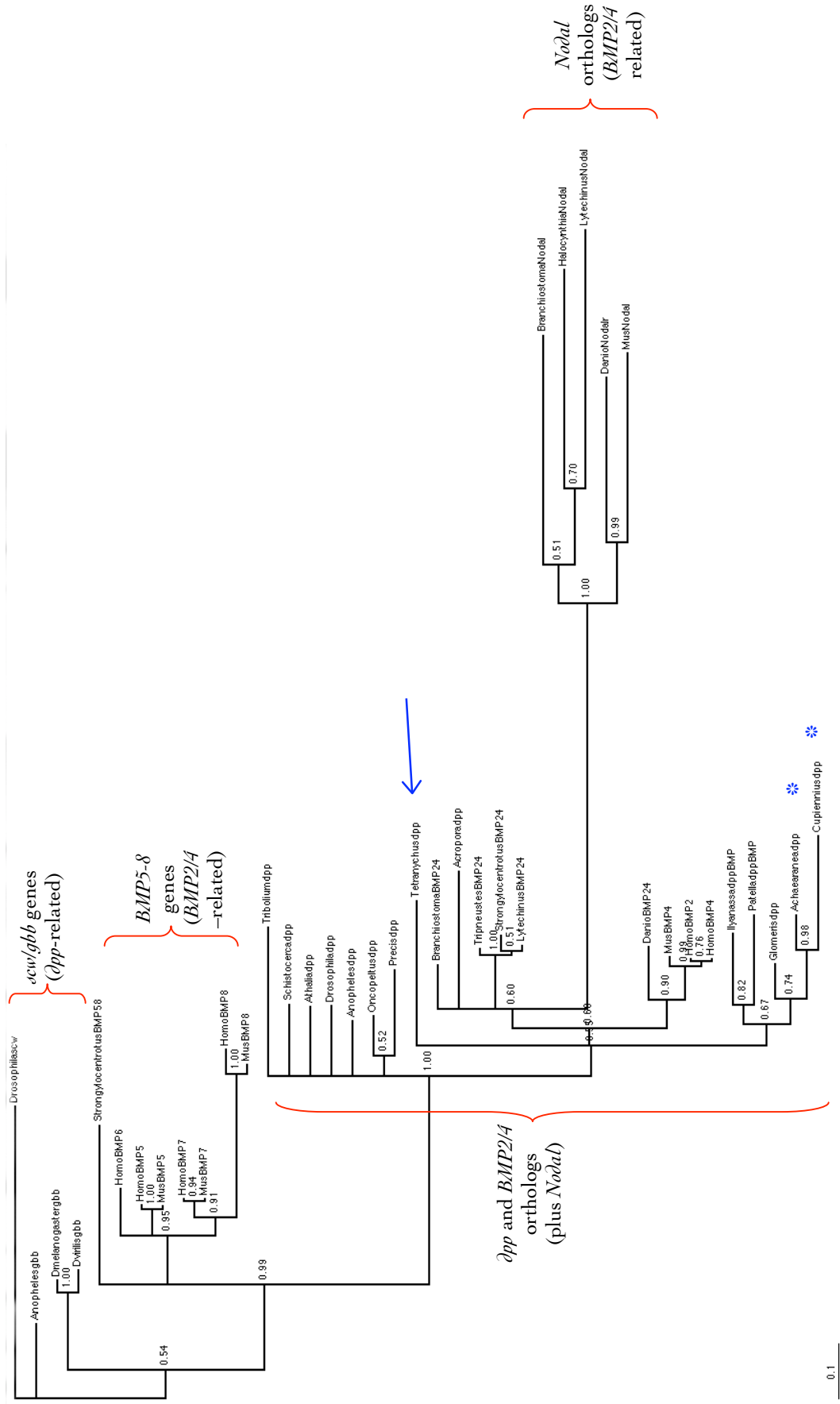


Figure 5.1.2c Bayesian consensus topology from analysis of *Tu-dpp* CDS (384bp) with *dpp/BMP2-4* and *dpp/BMP2-4*-related orthologs (*gbb*, *acw*, *Ndl*, *BMP5/6/7/8*) in a range of other taxa. *Tetranychus dpp* (blue arrow) branches within a fully supported (1.00 posterior probability) *dpp/BMP2-4* clade, indicative of true orthology. *Nodal* genes form a long branch clade, explaining anomalous placement within a sub-set of *dpp/BMP2-4* orthologs. Scale bar: expected base changes per site. Blue asterisks: other chelicerate *dpp* genes.

5.1.3 *Tu- ∂pp* ssRNA probe synthesis

Tu- ∂pp plasmid preparation

Having obtained the whole CDS for *Tu- ∂pp* , I designed gene-specific primers and amplified different length gene fragments by standard PCR as raw material for mRNA probe synthesis. Figure 5.1.3a illustrates the sequence and structure of the *Tu- ∂pp* gene, including target locations for ∂pp -specific primers, restriction enzyme (RE) cut sites - to be avoided in subsequent probe synthesis reactions - and predicted PCR product lengths. *Tu- ∂pp* DNA fragments were cloned into pGEM®-T Easy vector, enabling confirmation of insert size (and an approximate assessment of DNA concentration) prior to further probe synthesis, *via* a simple EcoRI restriction enzyme digest and agarose gel run (Figure 5.1.3a,iv). With respect to the ~1Kb *Tu- ∂pp* inserts in pBK-CMV vector plasmid, NotI/EcoRI double digests were necessary to confirm correct insert size, due to the nature of the multiple cloning site for this vector (Figure 5.1.3b,i-ii).

ssRNA probe synthesis reactions

Once correct insert size had been verified, single stranded, DIG-AP labelled RNA probes were synthesised from ∂pp inserts in pGEM®-T Easy and pBK-CMV vector. Single RE digests for plasmid linearization, and specific RNA polymerases for transcription, were selected as appropriate to (i) the particular vector plasmid, (ii) insert orientation in its vector and (iii) which mRNA probe polarity, sense or anti-sense, was desired from the transcriptional synthesis reaction (Figure 5.1.3b,iii).

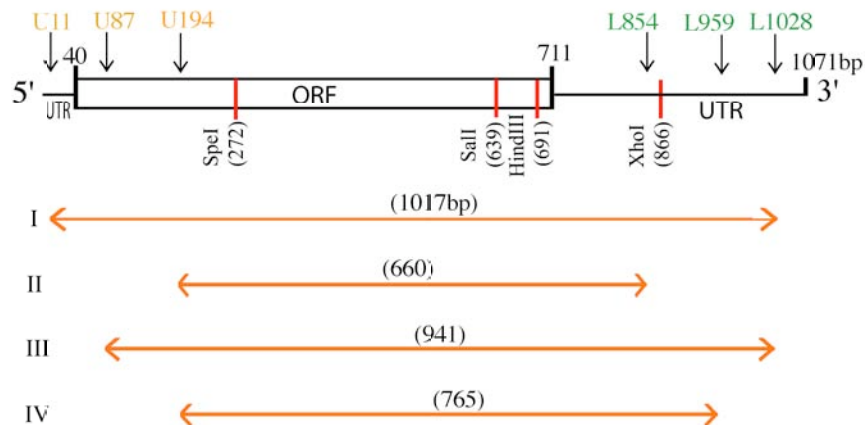
Confirmation of DIG-labelling

I used dot-blot tests to confirm effective α -DIG-AP labelling of probe RNA, cross-linking serially diluted samples of transcribed RNA to nitrocellulose membranes and detecting for phosphatase activity (further details in section 8.4.2 of Chapter VIII: Materials and Methods). For sense, antisense and control probes derived from inserts in both pGEM®-T Easy and pBK-CMV, decreased staining intensity is correlated with increased dilution factor, as would be expected according to associated exponential decline in target molarity (Figure 5.1.3b,iv).

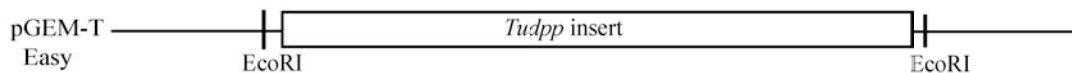
(i)

(1) CCGGCACGAGGTTTACTAACAGCAATATCATCTTGATGAGTGACCATGACTTTTCAA
 TGGACCCTTGTGGCAACAATTCTCTACATTGAGATTGCGGTTTGTCCTCCGTTTCATCGT
 ATCTGGTCACCATGACGATTCAAATCAACAGATTCTAATTTATTTAACCAATATCATC
 TGAATTGGACGAGTCTGGCTCTTGACGATCCAAGCCAATATTTTACACCACCAATTTT
 AAAACTGAGGAGAATGAACCTTTGCTAGAAGCACCTTTACTAGTGACTTATTCAAATGA
 TGGCACACCTGATGCCTCAAATGTTTTAAACGTGATAAGAGACAAGTTGATCAAAGAA
 ATAAGCCAGTCAGTGGACGTAAGAGTCGTCGCAACAAAGGAAAGAGGCATGTCGCCGG
 AGACCCATGTACGTCGATTTT**AGTGAAGTTGGTTGGAACAATTGGATTGTAGCTCCACT**
TGGTTATCAAGCATTTTATTGTGCGGGTGACTGTCCTCATCTCCTCAACGATGTCCATA
ATGCCACAAATCACGCCATTGTTCAAATTTAGTTAACTCTGTTACCCCTAACAAGGT
 CAAAAGGCTTGCTGTGTTCCCACCGAGCTTTCATCAATTTCAATTCTTTATGTCGACGA
 ATATGGAAAAGTTGTTCTCAAAAATTATCAAGACATGGTAGTGGAAGCTTGTGGCTGCC
 GGTAATCAACTATTGACAAAATCACTGCCAATTCACAACAAAACATCAGCCAAAGGTGT
 AAAATATCAATTGTAAAATTGTAAACACAAATGGAGCCTCTTATCTAGGCCTTCAAGAT
 TATAGTCAATTGCCATAAATGCTTTGTTTGGCAATTTCTACTCGAGTTTGAACAAC
 CAATTTTAATCAATTAGCAAACAAAACTATTTTACTTTAATTATTCTGATCGCTCTTG
 AAATCGAGTATCAACCAAAACCAACACA**AAAGTAAAAACAAAGCAAGGTTTGTGTAAT**
 TATAAATTGAATAAAAAACAAAAGGTAAACCAATTAACACATAAACAAAGAAAAA
 AAAAAAAAAA (1071)

(ii)



(iii)



(iv)

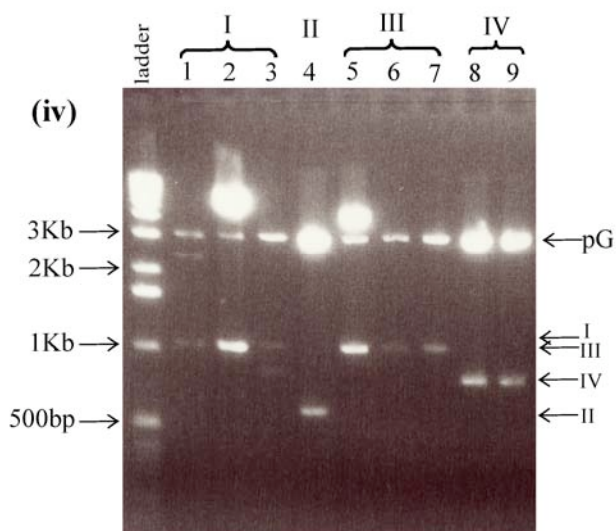
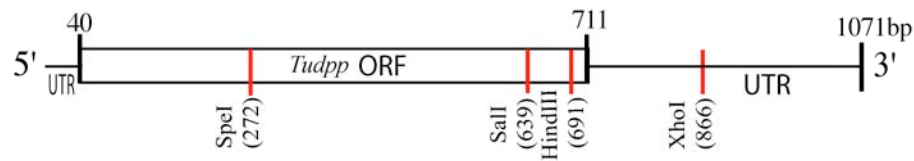
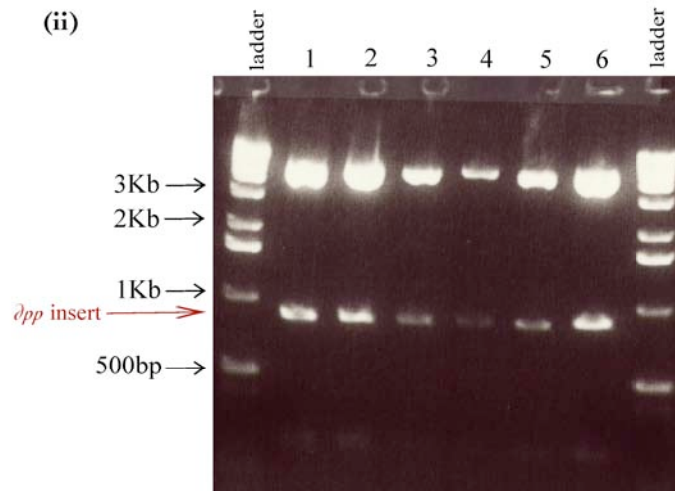


Figure 5.1.3a (i) *Tudpp* total sequence obtained; CDS in black, library screening probe in bold, primers used to re-amplify ISH probe material in yellow (U11, U87 & U194) and green (L854, L959, L1028). (ii) Schematic illustration of gene structure and fragment sizes (I - IV) obtained in PCR amplifying molecular probe DNA. RE cut sites also shown, relevant for insert size confirmation and plasmid linearisation reactions. (iii) Schematic: ISH probe fragments cloned into pGEM-T Easy vector, showing RE cut sites for insert size confirmation and DNA concentration assessment. (iv) EcoRI digested plasmid plus excised *dpp* inserts, run on 1% agarose gel. pG = pGEM-T-Easy vector without insert, I to IV = excised *dpp* inserts.

(i)



(ii)

**Figure 5.1.3b** *Tudp* ssRNA probe synthesis.

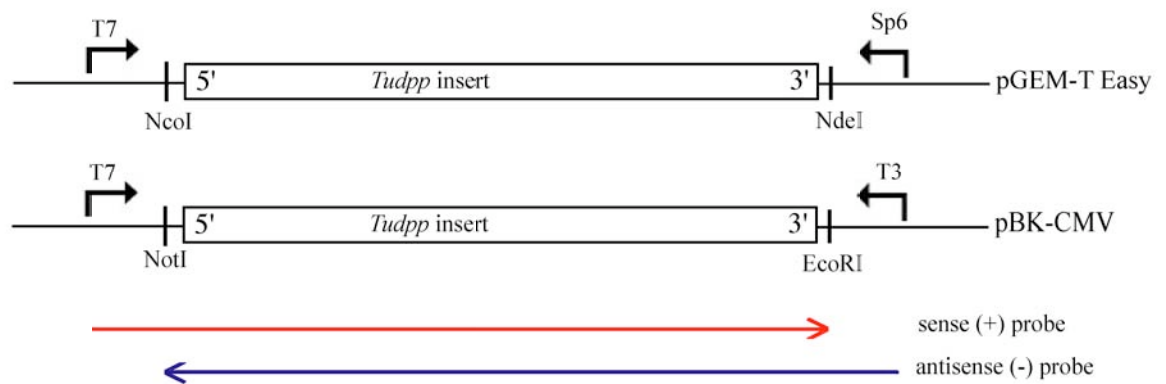
(i) Scaled schematic of sequenced *Tudp* gene, with cut sites to be avoided in probe synthesis.

(ii) Double digest (NotI/EcoRI) to excise *dpp* insert from pBK-CMV plasmid, for insert size confirmation. Run on 1% agarose gel.

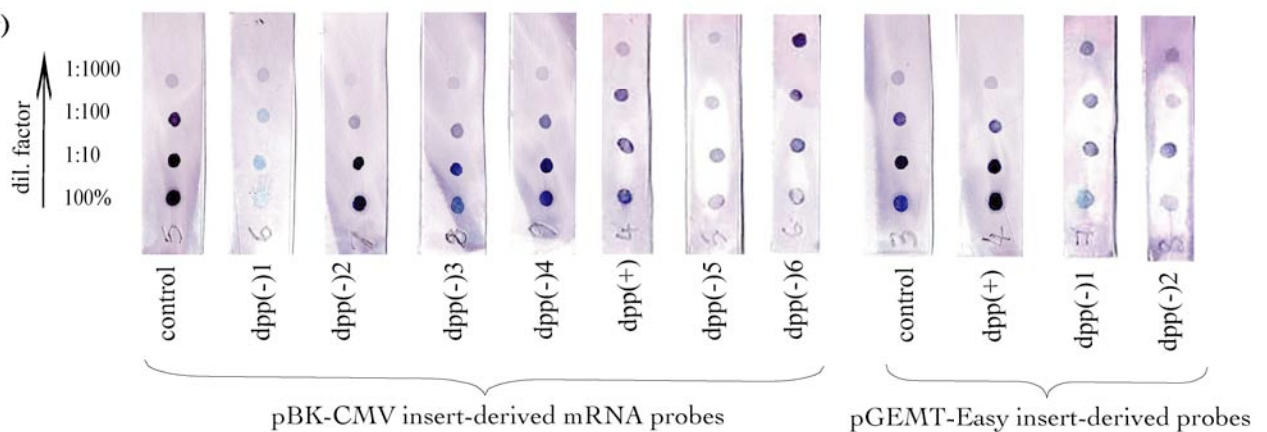
(iii) Schematics of *Tudp* inserts in pGEM-T Easy or pBK-CMV vectors. REs for plasmid linearisation are shown with relevant ssRNA polymerases for sense (red) and antisense (blue) probe transcription from 5'-3' oriented inserts.

(iv) Nitrocellulose membrane-bound RNA dot blot tests for efficient DIG labelling of probe.

(iii)



(iv)



5.2 *Tu- ∂pp* mRNA expression during embryogenesis

5.2.1 *Tu- ∂pp* mRNA detection: negative results

In situ hybridisation against *Tu- ∂pp* in blastoderm, germdisc, germband and late limb stage embryos was carried out using a number of ssRNA probes synthesised from inserts in either pGEM®-T Easy or pBK-CMV vector plasmids. I observed no specific hybridisation signal, save for one blastoderm stage embryo (hybridised with probe dppB3) in which positively stained blastomeres seemed to be located in only one hemisphere (Figure 5.2.1,ii).

Low level background ‘noise’ was apparent from late germband onwards, present at higher intensity in the more cell-dense late limb, ventral ridge and contracted germband tissue (Figure 5.2.1). The ubiquitous or diffuse background staining is indicative of non-specific probe binding or significant side-reactions generated during probe detection with phosphatase.

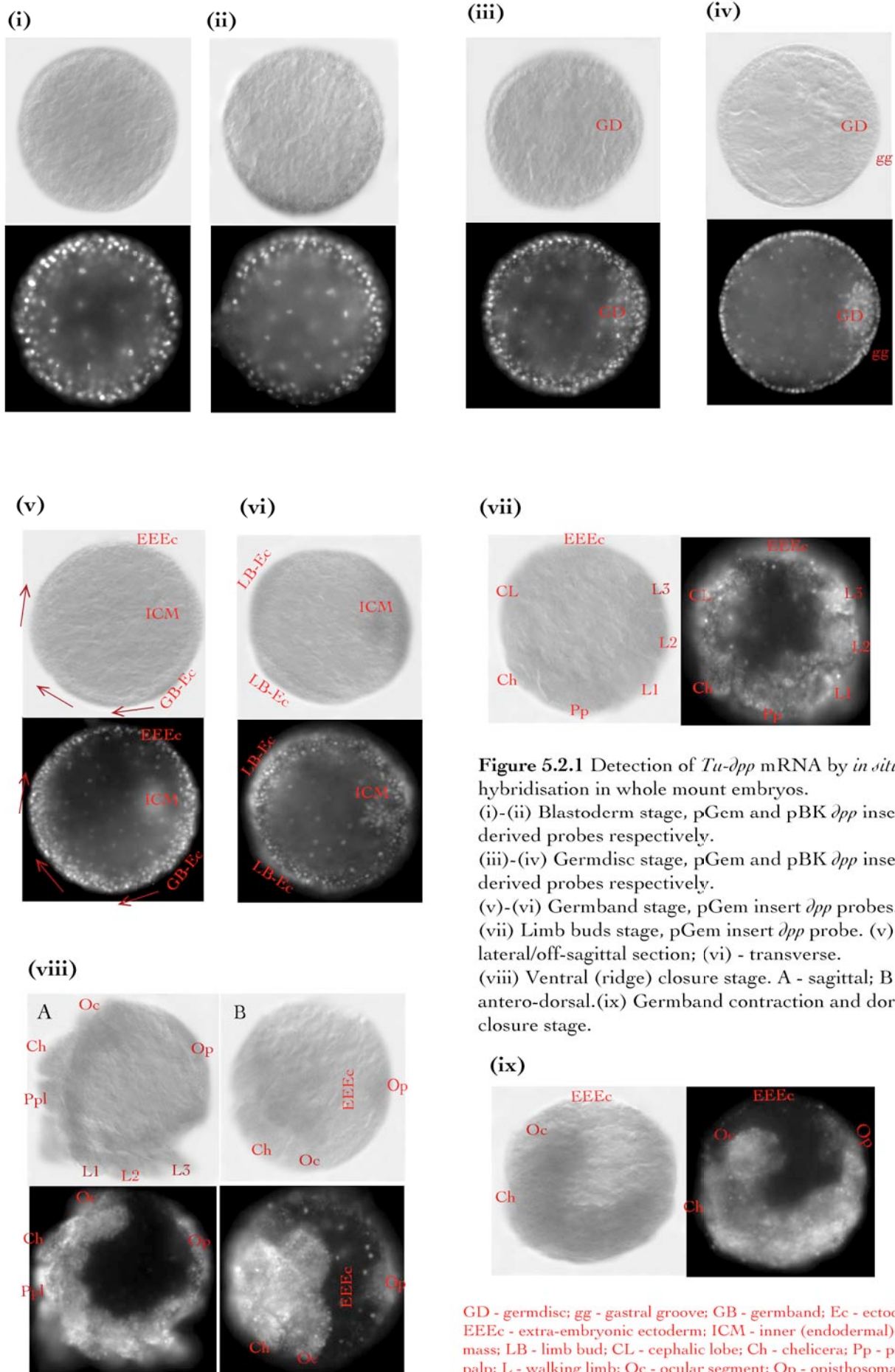
In view of negative results at all stages examined, conclusions as to the pattern of *Tu- ∂pp* transcription could not be made. It appears that for possible reasons outlined in section 5.2.2, probes failed to cross-hybridise with *Tu- ∂pp* mRNA *in situ*. Given the importance of Dpp signalling in metazoan development, lack of ∂pp -specific hybridisation throughout spider mite embryogenesis emphasises the probability of failed mRNA-probe binding, rather than demonstrating absence of *Tu- ∂pp* transcription *per se*.

5.2.2 Reasons for negative results

Failed hybridisation of single stranded ∂pp mRNA probes to target *Tu- ∂pp* mRNA in whole mount embryos could have the following explanations, each subsequently criticised:

1. Degradation of target mRNA during embryo fixation.
2. Degradation of ssRNA probe prior to hybridisation.
3. Failure of probe molecules to penetrate the embryo effectively.
4. Lack of *Tu- ∂pp* transcription at the stage investigated.

Degradation of *Tu- ∂pp* mRNA during embryo fixation (explanation 1.) is not impossible, especially as RNA is fragile, but it is made improbable given that *in situ* hybridisation for other genes worked after the same fixation protocol. Probes could have degraded to a level that prevented bonding with complementary target RNA (explanation 2.). Although some other mRNA probes handled similarly in between synthesis and hybridisation hybridised successfully, it could be that the ∂pp probe RNA



was particularly susceptible due to uncharacterised molecular instability, or was exposed unknowingly to destructive environmental RNase activity, in spite of precautions that were taken to avoid this. I tried *in situ* hybridisation with multiple *∂pp* probes between ~650 - 1Kb in size, i.e. at molecular weights unlikely to impede cell penetration (explanation 3.), especially considering that successful *in situ* results were obtained for other genes using probes within the same molecular size range.

Lack of any *Tu-∂pp* transcription during embryogenesis (explanation 4.) would be surprising given the important role of TGF- β signalling in animal development. Roles in dorso-ventral axis specification, later boundary formation and/or morphogenetic events (e.g. ecto- vs. endoderm differentiation, proximo-distal limb proliferation/repair, neural patterning) are conserved in many arthropods and other metazoans as disparate as molluscs, corals, sea anemones, crinoids and chordates (Akiyama-Oda and Oda 2003; Darras and Nishida 2001; Finnerty et al. 2004; Goto and Hayashi 1997; Hayward et al. 2002; Holland 2004; Irish and Gelbart 1987; Jockusch and Ober 2004; Nederbragt et al. 2002; Padgett et al. 1993; Patruno et al. 2002; Prpic 2004a; Prpic et al. 2005; Serrano and O'Farrell 1997; St Johnston and Gelbart 1987). In addition, a *Tu-∂pp* EST transcript was recently reported from a *Tetranychus urticae* EST project at Agriculture Canada (A. Khila, pers. com.), so the mRNA must be transcribed at some point.

5.3 Detection of *∂pp* signal transduction in *Tetranychus urticae* and *Cupiennius salei* embryos via a phospho-Smad-1 antibody

5.3.1 Rationale for detecting Dpp activity via phospho-Smad-1

Dpp ligand binding and activation of TGF- β transmembrane receptors results in phosphorylation of the cytokine Smad-1, at 2 serines in the Smad-1 isoform-specific C' motif SSXS. This phosphorylation causes dimerisation with the isoform Smad-4, followed by nuclear translocation and DNA binding to activate target gene transcription, such that p-Smad-1 distribution precisely mirrors Dpp signal transduction patterns (Gilbert 2000; Lewin 2002; Persson et al. 1998; Suga et al. 1999). An antibody to the C-terminal peptide of phosphorylated Smad-1 (p-Smad-1) is available, and has been shown to cross-react in *Drosophila melanogaster* and the spider *Achaearanea tepidariorum* (Akiyama-

Oda and Oda 2003; Kretzschmar et al. 1999; Persson et al. 1998). Given this example of epitope conservation between a basal arthropod (spider) and pancrustacean (fly), there is a strong possibility of p-Smad-1 antibody cross-reactivity in chelicerates other than *Achaearanea*. Therefore, I chose to test the feasibility of using p-Smad-1 antibody¹⁷ to visualise Dpp signal transduction patterns in whole mount *Tetranychus urticae* embryos. Embryos of the primitive spider *Cupiennius salei* were also available to me, and as *Co-dpp* mRNA transcript localisation is documented in Prpic (2003), I tested the p-Smad-1 antibody in this species, aiming to complement the available *dpp* mRNA data with actual signal transduction patterns (Prpic et al. 2003).

5.3.2 *Tetranychus urticae* p-Smad-1 distribution

i) Blastoderm stage

In experimental embryos, cells stained positively for DIG-AP labelled p-Smad-1 protein localise predominantly in one hemisphere of the blastoderm, the polarity of which is unclear due to its homogeneous structure before ~10hrs AEL (Figure 5.3.2a,iii-iv). Appearing polarised, Dpp signal transduction may be providing early axial or positional information, interpreted for example by breaking symmetry and restricting cell proliferation to the proto-germdisc side. Asymmetric activation of Dpp/BMP2/4 signalling has been demonstrated to play a role in secondary, most often dorso-ventral axis specification throughout the Bilateria, and Dpp transduction on one side of a pre-germdisc acarid embryo, if borne out by further data, would provide another example of early Dpp-mediated D/V axis specification mechanisms in arthropods (Akiyama-Oda and Oda 2003; Akiyama-Oda and Oda 2006; De Robertis and Sasai 1996; Finnerty et al. 2004; Hayward et al. 2002; Holland 2004). However, a degree of background staining and diffuse signal, as well as the effects of partly quenched DAPI fluorescence in the large blastoderm cells, make it difficult to commit conclusively to a hemispherical p-Smad-1 domain at this stage: artefactual staining, or binding of α -p-Smad-1 to non-specific target(s), are alternative explanations for the putative pattern observed. Control blastoderm embryos (Figure 5.3.2a,i-ii) were successfully stained for α -phospho-HistoneIII, a mitotic cell centrosome marker.

¹⁷ The antibody distribution patterns described reflect hydrogen peroxide/H₂O₂ reduction by peroxidase and precipitation of oxidised digoxigenin, derived from 2° α -rabbit DIG-AP conjugated to 1° α -pSmad-1.

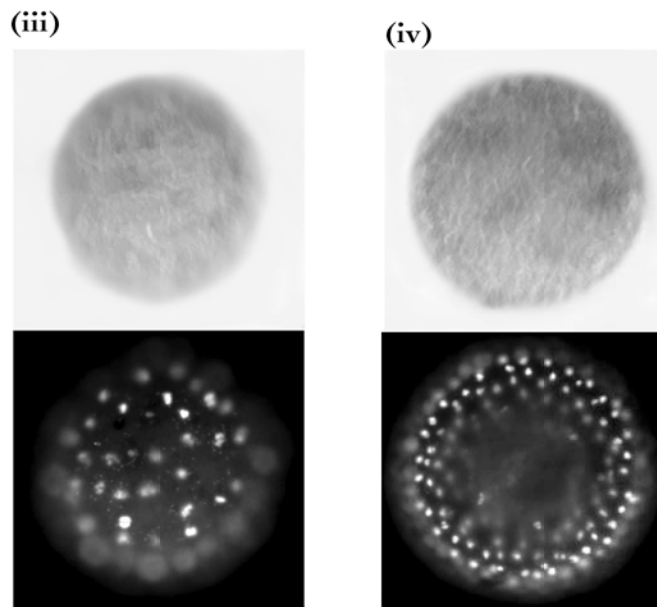
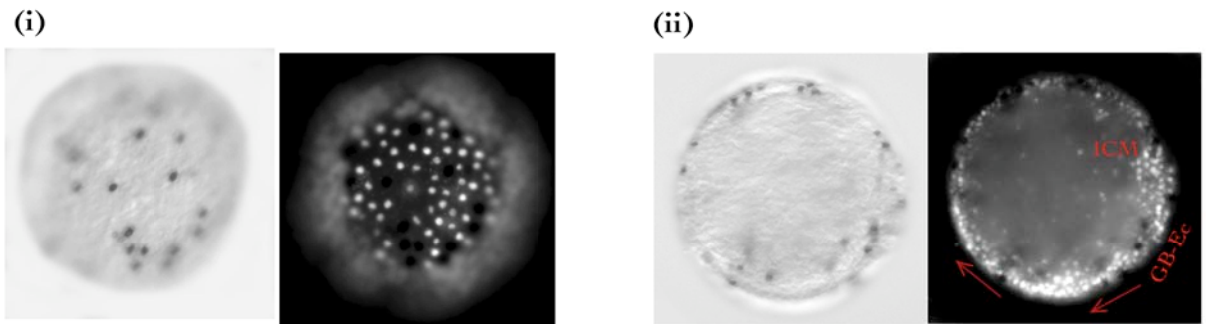
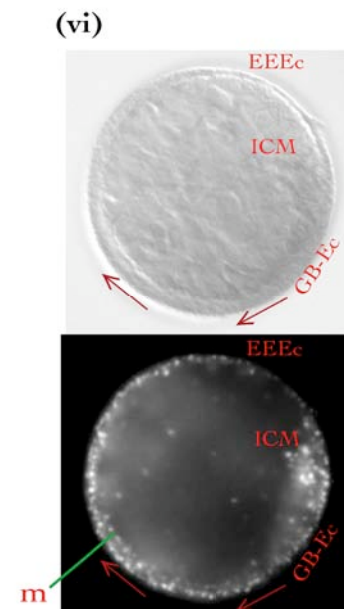
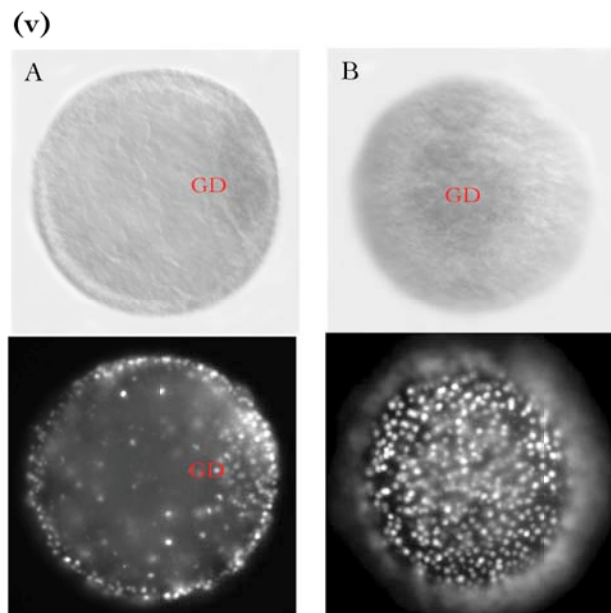


Figure 5.3.2a Anti-pSmad1 distribution. (i)-(ii) Anti-pHistoneIII positive control: blastoderm (i) and germband (ii) stage. (iii) Early blastoderm pSmad1 antibody detection: hemispherical (upper) domain? (iv) Late blastoderm pSmad1: possible hemispherical (upper) staining? (v) Germdisc: A - sagittal and B - lateral, just beneath disc surface. Artefact of anti-pSmad1 'detection' in more cell-dense GD. (vi) Germband, sagittal section. pSmad1-antibody again detected in more dense, proliferating germband ectoderm cells. Anterior left, Dorsal up in germdisc-band stage embryos.

GD - germdisc; EEEc - extra-embryonic ectoderm; ICM - inner cell mass; GB-Ec - germband ectoderm; m - mesoderm



ii) Germdisc stage

Background staining was higher in experimental than control (α -pHistone-III) embryos, but signal:noise conditions were good enough to observe apparent p-Smad-1 localisation to the inner germdisc cell mass that underlies future posterior germband ectoderm (Figure 5.3.2a,v). This protein distribution may indicate:

- (i) short-range Dpp signal transduction induced by Dpp ligand activity from overlying ectoderm (juxtacrine signalling), or
- (ii) short-range Dpp transduction induced by ligand activity within the inner/proto-endodermal cell lineage itself (auto/juxtacrine signalling), or
- (iii) α -p-Smad-1 binding to non-specific, non-target epitopes in the inner cell mass environment.

A range of trafficking control mechanisms allow Dpp signal transduction to occur at short or long range, enabling it to act as a morphogen in fields encompassing whole D/V axes down to single distal limb digits (Drossopolou et al. 2000; Wang and Ferguson 2005). Dpp ligands may be held in an inactive 'latent complex', and once secreted and dimerised may be bound by extracellular matrix/cell surface heparan sulfate proteoglycans (HSPGs) such as Dally, degraded in the extracellular space, titrated by cell surface receptors or subject to planar transcytosis (endocytic trafficking) mediated by Dynamin (Entchev et al. 2000; Fujise et al. 2003; Tabata and Y. 2004). Better antibody staining results, *dpp* mRNA transcript and Dpp protein localisation would be required to distinguish the possibility of Dpp trafficking and signal transduction dynamics from, to or within, the inner/endo-mesodermal cell mass from the possibility of non-specific p-Smad-1 antibody binding.

iii) Germband stage

Germband elongation is accompanied by detection of limited Smad-1 phosphorylation in migrating mesoderm cells and the distinct inner cell mass that underlies the presumptive opisthosoma, with low levels just detectable in the overlying germband ectoderm (Figure 5.3.2a,vi). The diffuse and inconsistent nature of this antibody staining pattern makes it tentative, and not a conclusive report of p-Smad-1 activity and hence Dpp signal transduction in the germband: non-target epitope binding or background staining effects may be additionally, or exclusively, at play.

If proved genuine, Dpp signal transduction in the *Tetranychus* germband mesoderm could suggest some involvement in mesoderm proliferation and/or directing migration from the posterior growth zone. Considering genetic studies in spiders *Cupiennius*, *Achaearanea* and *Pholcus*, this mesodermal role would be novel within chelicerates, yet remains consistent with TGF- β function in divergent metazoan developmental processes involved with proliferation and morphogenesis (Akiyama-Oda

and Oda 2006; Hayward et al. 2002; Patruno et al. 2002; Prpic et al. 2003). High level mesodermal and low level ectodermal transduction may result from Dpp diffusion from mesoderm to ectodermal cells down a transient gradient, or trafficking of ectodermal Dpp to the mesoderm by directing Dpp to basal cell membranes, adjacent to mesoderm (Entchev et al. 2000; Persson et al. 1998). Such an interaction between germ layers would emphasise flexibility in TGF- β signalling dynamics, being the reverse of that observed in the spider *Achaearanea*, where cumulus mesenchyme cells signal to overlying ectoderm inducing dorso-ventral axis specification (Akiyama-Oda and Oda 2003).

iv) Limb growth - germband contraction stages

From early limb development to the stage of ventral ridge closure, I detected clear, consistent p-Smad-1 antibody binding in association with the inner cell mass, a putative endodermal cell cluster ~10 μ m in diameter, located beneath and at least one cell diameter separate from, the posterior opisthosomal ectoderm (Figure 5.3.2b,iii-iv). From ventral closure onwards, an additional patch of weaker p-Smad-1 staining was evident within the cephalic midline (Figure 5.3.2b,v-vi). Positive and negative control embryos were as expected, and within experimental embryos I observed almost no background noise, nor any evidence for Smad-1 phosphorylation anywhere but the cephalic and sub-opisthosomal patches.

Lack of Dpp signalling in the developing spider mite limb is surprising, given that *dpp* has been implicated with a role in proximo-distal limb patterning in a wide range of arthropods, from insects to arachnids (Akiyama-Oda and Oda 2003; Akiyama-Oda and Oda 2006; Jockusch et al. 2000; Jockusch and Ober 2004; Niwa et al. 2000; Prpic et al. 2003; Sanchez-Salazar et al. 1996). Indeed, the restriction of apparent p-Smad-1 activity to anterior and posterior cell clusters could be an erroneous report of Dpp transduction derived from non-target epitope binding, or from a simple physical property of the cells themselves: for example, the sub-opisthosomal and later central cephalic domains are compellingly linked – both spatially and temporally - to independent invagination of the proctodeum and stomodeum respectively (c.f. Chapter II) (Anderson 1973).

5.3.3 *Cupiennius salei* α -p-Smad-1 distribution

i) Germdisc stage

In the late germdisc stage of *Cupiennius salei*, the p-Smad-1 antibody is most strongly detected in a semi-lunar domain that encompasses much of the proliferating germdisc (Figure 5.3.3a,i). The level of presumed Dpp signalling appears higher at the posterior growth zone margin of this germdisc

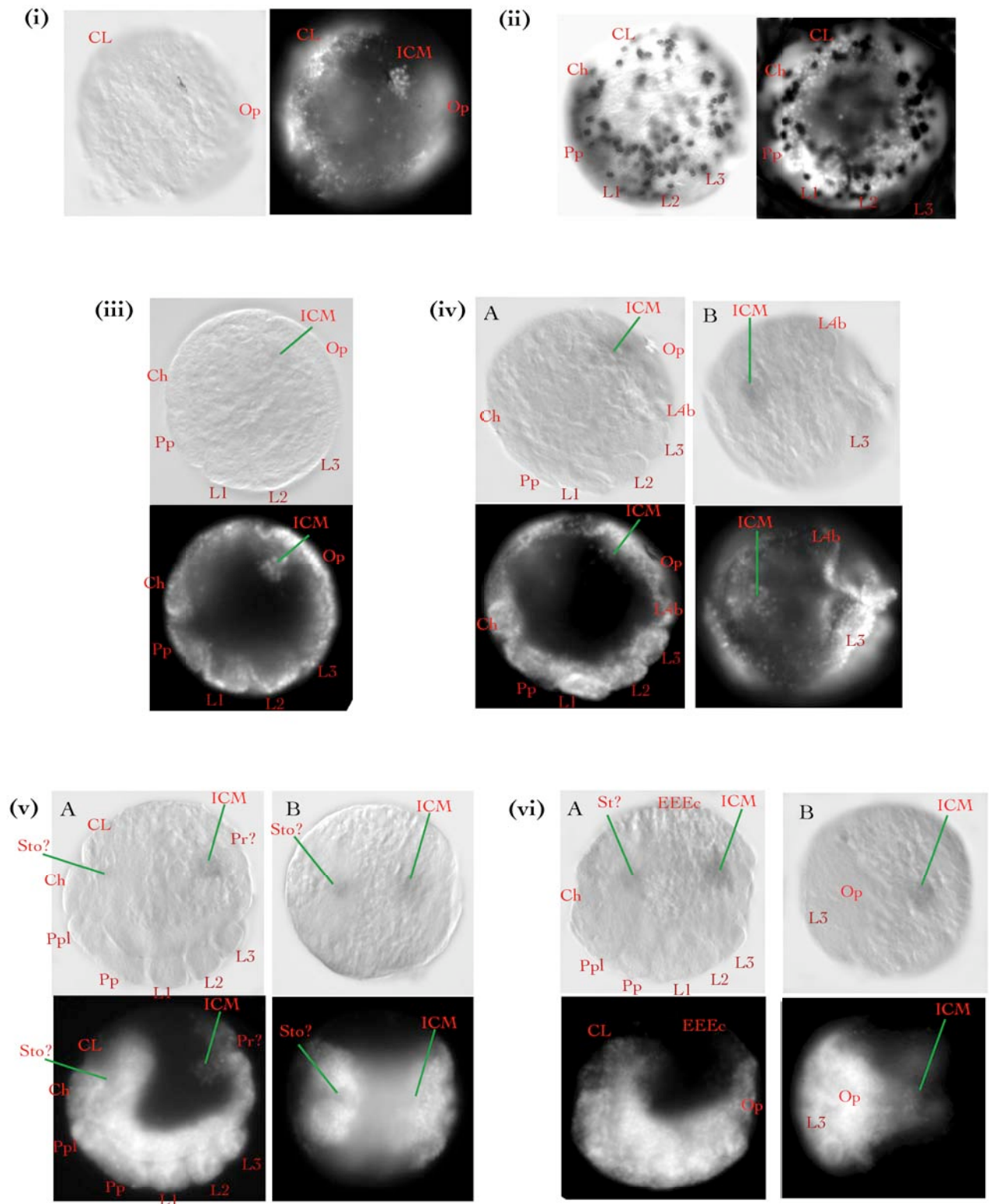


Figure 5.3.2b Detection of Smad1 phosphorylation with anti-pSmad1. (i) Negative control: germband stage. (ii) Anti-pHistone-III positive control: limb bud stage. (iii) Mid limb bud stage: no pSmad1 activity detected. (iv) Ventral ridge closure stage: antibody localisation at inner cell mass, beneath opisthosoma - artefact? A - sagittal; B - postero-ventral. (v) Ventral closure/GB contraction stage: pSmad1 detected in two patches, one at inner cell mass, one within medial cephalic lobe - artefact related to proctodeum/stomodeum invagination? A - sagittal; B - dorsal transverse. (vi) Germband contraction/dorsal closure stage: pSmad1 in 2 patches as in (v). A - sagittal; B - posterior. Anterior left.

CL - cephalic lobe; Op - opisthosoma; ICM - inner (sub-opisthosomal) cell mass; Ch - chelicera; Pp - pedipalp; Ppl - Pp lobe; L1 - walking limb pair 1; St? - putative stomodeum invagination region; Pr? - putative proctodeum invagination region; D - germdisc; EEEc - extra-embryonic ectoderm

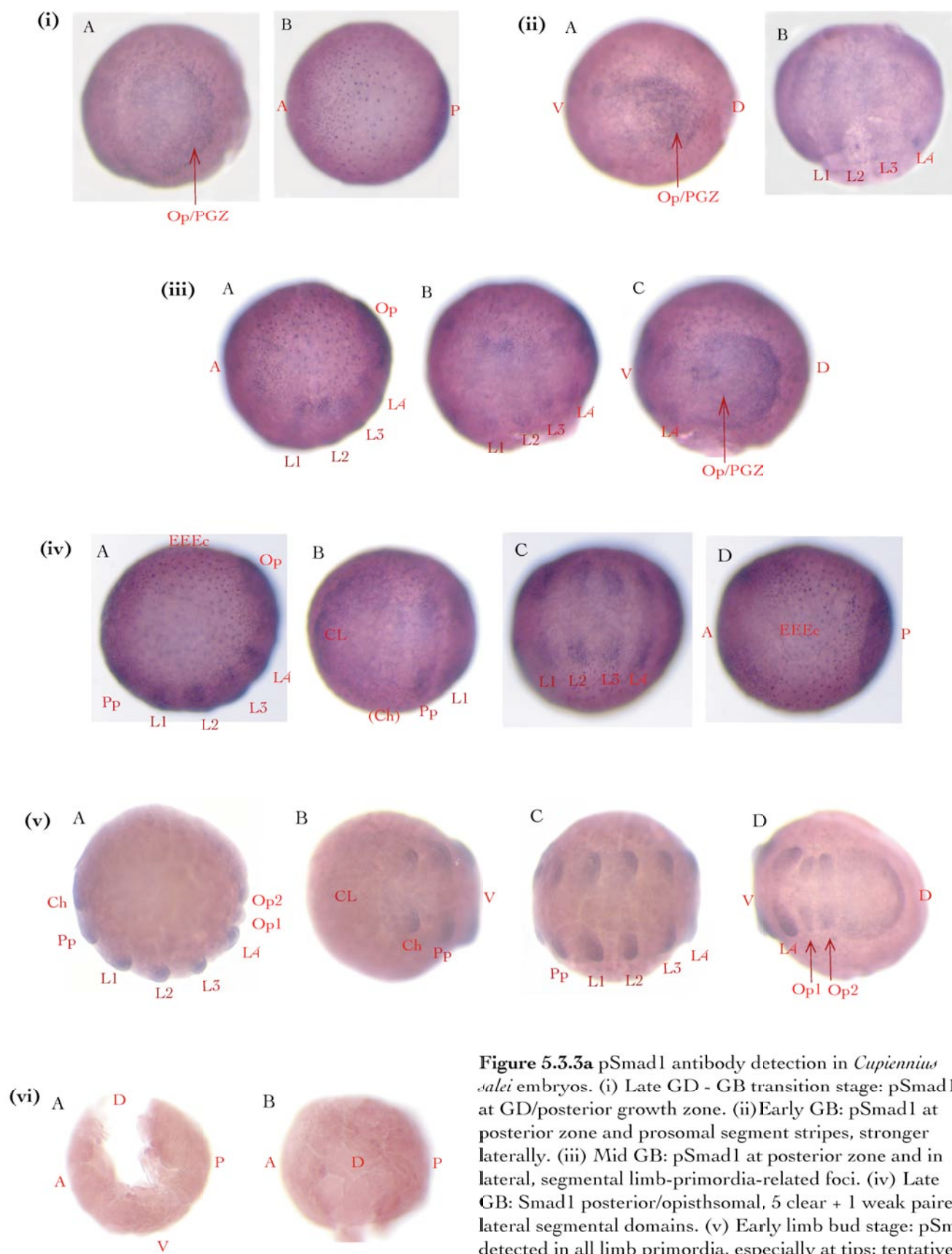


Figure 5.3.3a pSmad1 antibody detection in *Cupiennius salei* embryos. (i) Late GD - GB transition stage: pSmad1 at GD/posterior growth zone. (ii) Early GB: pSmad1 at posterior zone and prosomal segment stripes, stronger laterally. (iii) Mid GB: pSmad1 at posterior zone and in lateral, segmental limb-primordia-related foci. (iv) Late GB: Smad1 posterior/opisthosomal, 5 clear + 1 weak paired lateral segmental domains. (v) Early limb bud stage: pSm1 detected in all limb primordia, especially at tips; tentatively at posterior opisthosomal margin. (vi) Negative control.

EEEc - extra-embryonic, dorsal ectoderm; CL - cephalic lobe; Op/PGZ - opisthosoma/posterior growth zone; Op - opisthosomal segment; L - walking limb pair; Pp - pedipalp; Ch - chelicera; D - dorsal; V - ventral; A - anterior; P - posterior; GB - germband; GD - germdisc.

domain (Figure 5.2.2.1, iiD), suggesting a possible general role for Dpp in promoting germdisc cell proliferation. However, the antibody could be merely localising within this region of high cell density and physical complexity, conclusive interpretation being further limited by the presence of background staining and non-specific staining of large, extra-embryonic cells at this stage.

ii) Germband stages

Early germband

In the early germband, α -p-Smad-1 reports Smad-1 phosphorylation within:

1. the telson/opisthosomal rudiment that arises from posterior germdisc cells, particularly at lateral and posterior margins, and
2. diffuse, hemisegmental stripes that fade at the ventral midline (Figure 5.3.3a,ii).

The posterior p-Smad-1 domain persists from the germdisc stage, and may derive from cell density effects, Dpp proliferation signals or, as it is strongest at the lateral and posterior edges of the growth zone, some kind of dorsal specification. A function for ∂pp in dorsal identity of the posterior germband could represent partial conservation relative to other arthropods; for example, ∂pp marks the dorsal (peripheral) germband edges in the coleopteran *Tribolium castaneum* and orthopterans *Gryllus bimaculatus*, *Schistocerca gregaria* and *S. americana* (Dearden and Akam 2001; Jockusch et al. 2000; Niwa et al. 2000; Sanchez-Salazar et al. 1996). The striped, segmental pattern mirrors that published for *Cupiennius salei* ∂pp mRNA expression, and hence indicates short-range or autonomous Dpp signalling during early segmentation, just prior to definition of limb primordia (Prpic et al. 2003).

Mid germband

The p-Smad-1-positive 'stripe' domains become restricted ventrally to form bilateral pairs of cell clusters in each segment, associated with prosomal limb primordia prior to any outgrowth (Figure 5.3.3a,iii). Cell density is relatively low within the bilateral clusters, so a genuine report of active Dpp signalling in limb primordia seems more likely than background staining or cell density effects. In most clusters, antibody-marked boundaries are difficult to discern and hence the position of each parasegment difficult to infer – specifically in relation to *Cs- ∂pp* or *At- ∂pp* expression, in which a linear posterior ∂pp transcription boundary abuts anterior parasegment borders (Akiyama-Oda and Oda 2003; Prpic et al. 2003).

Apparent α -p-Smad-1 signal persists in the opisthosomal region, especially at the posterior to lateral margin (Figure 5.3.3a,iiiD). As mentioned previously, this may or may not be a genuine report of Dpp signal transduction in presumptive dorsal tissues as higher cell density, coupled with ubiquitous background staining, makes conclusive interpretation challenging.

Late germband

α -p-Smad-1 patterns in the late germband, in which primordia for Ch-L4 prosomal appendages are apparent, is similar to but somewhat more intense than that reported for mid germband embryos, with similar problems arising from background staining interference and cell density effects. The p-Smad-1 antibody is detected throughout the limb primordial cells (Figure 5.3.3a,iv), with increased $[p\text{-Smad-1}]_{\text{nuclear}}$ possibly a manifestation of *dpp* up-regulation relative to previous stages. Posterior boundaries of segmental clusters of p-Smad-1 are more defined, perhaps related to more precise regulation at parasegment borders, or purely due to a stronger signal revealing more detail.

In addition to limb-related expression, α - p-Smad-1 continues to be detected in the presumptive opisthosomal region, and also at the anterior periphery of the accumulating cephalic lobe (Figure 5.3.3a,ivB). These anterior and posterior p-Smad1 domains may report Dpp signal transduction in presumptive dorsal tissue, or are artefacts created by high tissue density and background staining problems.

iii) Early limb development

Non-specific background signal is very low in limb bud stage embryos onwards, conferring more reliability to the observed distribution of *Cupiennius salei* p-Smad-1. Limb bud outgrowth (120-130hrs AEL) is concomitant with α -p-Smad-1 restriction to the distal tip of all prosomal, 1st and 2nd opisthosomal segment appendages (Figure 5.3.3a,v). A low level of Smad-1 phosphorylation characterises medial limb tissue, suggestive of down-regulated *dpp* transcription in non-distal cells. This pattern exactly matches that observed during limb bud outgrowth in *Achaearanea tepidariorum* and *Cupiennius salei*, as well as being reminiscent of distal appendage *dpp* transcription in millipede *Glomeris marginata* and insects *Thermobia domestica*, *Schistocerca gregaria*, *S. americana*, *Gryllus bimaculatus* and *Tribolium castaneum* (Akiyama-Oda and Oda 2006; Dearden and Akam 2001; Jockusch et al. 2000; Jockusch and Ober 2004; Niwa et al. 2000; Prpic 2004a; Prpic et al. 2003; Sanchez-Salazar et al. 1996). Distal restriction of the Dpp morphogen may be significant with respect to the nature of molecular interactions with factors such as Wingless within the antero-posterior axis of developing three dimensional appendages. The potential effects of signal topology during limb patterning are considered further in Chapter VII - Discussion, along with implications for plasticity in early *Distal-less* activation mechanisms, given recent functional data on the importance (or not) of a complete Wg/Dpp signalling module in arthropod limb allocation (Akiyama-Oda and Oda 2006; Angelini and Kaufman 2004; Hejnol and Scholtz 2004; Miyawaki et al. 2004; Prpic 2004a; Prpic et al. 2003; Yamamoto et al. 2004).

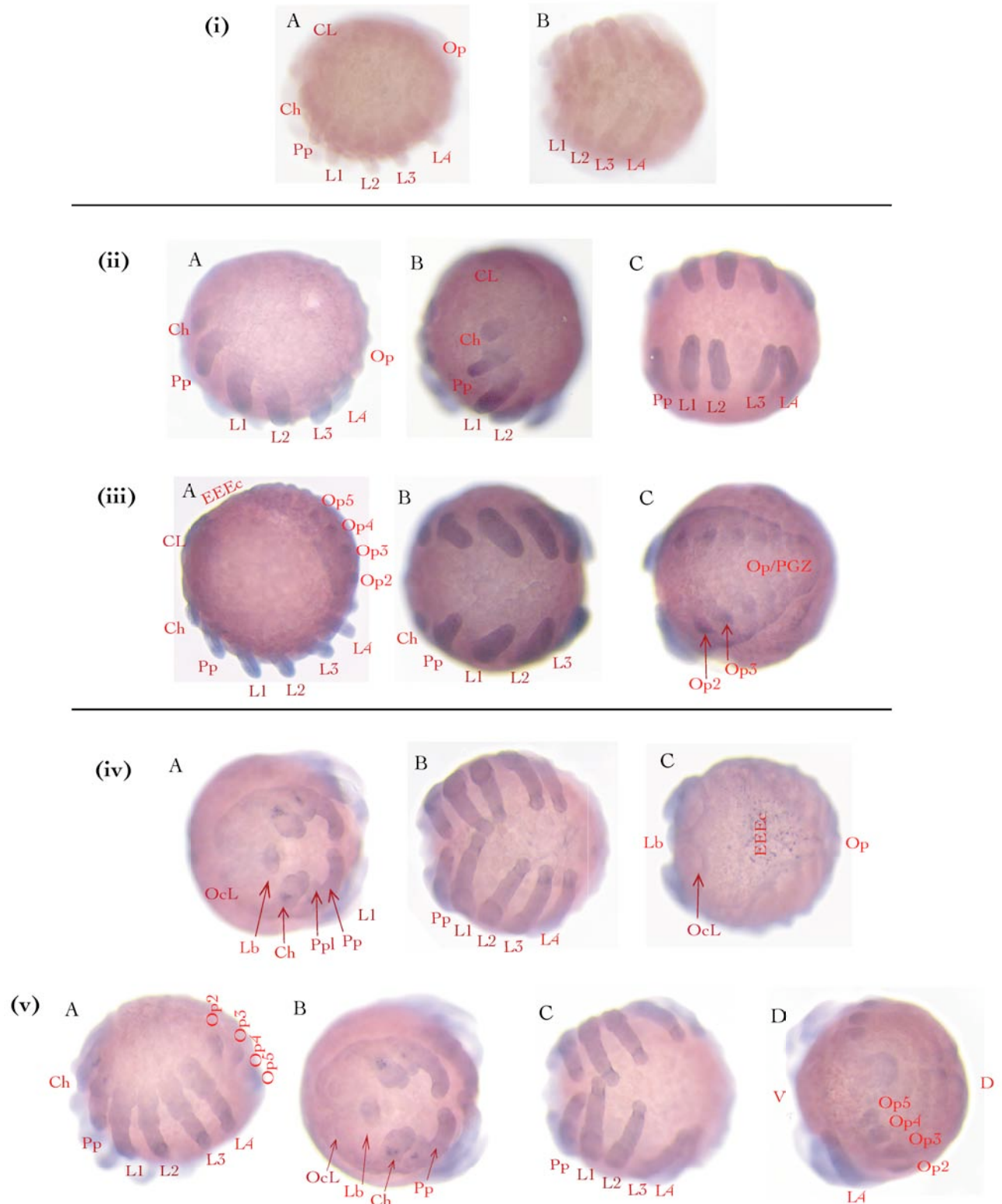


Figure 5.3.3b pSmad1 antibody detection during limb elongation and inversion in *Cupiennius salei* embryogenesis. (i) Negative control. (ii)-(iii) Early limb elongation, prior to inversion and complete opisthosomal segmentation. pSmad1 activity in throughout prosomal limbs and emerging Op rudiments. (iv)-(v) Inversion stage: pSmad1 detected strongly in proximal tarsal rings, cell clusters within chelicerae and Pp lobes, and at lower levels in all limb/rudiment tissue, including the labrum.

CL - cephalic lobe; OcL - ocular lobe; Lb - labrum; Ch - chelicera; Pp - pedipalp; Ppl - Pp lobe; L - walking limb; Op - opisthosoma; Op2 - book lung segment/appendage; Op3 - tubular tracheal segment/appendage; Op4/Op5 - spinneret segments/appendages; D - dorsal; V - ventral; A - anterior; P - posterior.

iv) Mid limb development

I detected α -p-Smad-1 in the distal part of extending limbs at early inversion/limb morphogenesis. By later inversion stages, ubiquitous low level p-Smad-1 activity is observed throughout all but the most proximal parts of prosomal appendages, stronger activity persisting distally (Figure 5.3.3b,ii-iii). Distally transduced Dpp signal is consistent with previous studies of genetic interactions during proximo-distal arthropod limb patterning, as is modulation of *ppp* transcription domains as limb morphogenesis proceeds (Akiyama-Oda and Oda 2006; Angelini and Kaufman 2004; Campbell et al. 1993; Dahn and Fallon 1999; Hejnowicz and Scholtz 2004; Jockusch et al. 2000; Jockusch and Ober 2004; Miyawaki et al. 2004; Penton and Hoffmann 1996; Prpic 2004a; Prpic et al. 2003; Sanicola et al. 1995; Yamamoto et al. 2004)

v) Late limb development: Dynamic and specific p-Smad-1 detection

p-Smad-1 in pedipalps and walking limbs

Concurrent with full inversion and dorsal closure, Pp-L4 appendages elongate and acquire clear morphological signs of jointing etc. The pattern of p-Smad-1 during late Pp-L4 limb development is dynamic, protein being expressed in a ring marking the distal podomere boundary, between the tarsus and metatarsus (see Figure 5.3.3b,iv-v). Ubiquitous low level expression in remaining limb tissue may derive from weaker *ppp* signalling distal to the trochanter (2nd podomere), but this conclusion is confounded by a degree of background staining within the limbs. With or without additional medial p-Smad-1 translation, evidence points to a dynamic role for Dpp signalling in proximo-distal appendage patterning in *Cupiennius salei*, perhaps particularly with reference to joint formation, segment allometry or specification of sensory organs that are predominantly associated with joints.

p-Smad-1 in chelicerae and Pp lobes

I observed strong p-Smad-1 antibody staining in a discrete cluster of ~4-6 cells located medially in the basal/coxal segment of chelicerae and pedipalpal lobes (Figure 5.3.3b,iv-v). The affinity of these cells is uncertain; they may relate to a specific sensory neuronal or endocrine structure (Anderson 1999; Foelix 1996). Proof of this affinity could imply a role for Dpp signalling in directing late chelicer and pedipalp lobe organogenesis. However, the intensity of chromophore precipitate in the cells also raises the possibility of a degree of non-specific staining, perhaps due to a physical structure (e.g. sensory terminus or gland duct opening) that preferentially traps antibody molecules.

5.3.4 Smad-1 phosphorylation: *Cupiennius* vs. *Tetranychus*

Comparison of α -p-Smad-1 data

The pattern of Dpp signal transduction reported *via* p-Smad-1 antibody detection in *Cupiennius salei* embryogenesis differs markedly from *Tetranychus urticae* p-Smad-1 distribution at comparative stages of development. Although both species appear to exhibit active Dpp signalling in the germdisc, transduction occurs in more posterior/dorsal ectoderm in the *Cupiennius salei* late germdisc, and in *Tetranychus urticae* primarily in ingressed mesoderm cells. During germband formation in *Cupiennius salei*, α -p-Smad-1 is detected in hemi-segmental stripe domains that contract ventrally to mark paired, lateral limb primordia prior to morphogenesis; in contrast, I only saw α -p-Smad-1 in *Tetranychus urticae* germband development within the sub-opisthosomal cell mass. Limb development is accompanied by distally localised then dynamic Dpp signal transduction in specific proximo-distal appendage domains in *Cupiennius salei*, whereas α -p-Smad-1 is only visible in *Tetranychus urticae* at the sub-opisthosomal cell mass and a medial cephalic domain.

Evolutionary implications regarding Dpp signals

Considering available *dpp* expression data for chelicerates, myriapods, crustaceans and insects, it appears that (i) posterior peripheral/dorsal germband, (ii) segmental limb primordia and (iii) dynamic appendage associated p-Smad-1 in *Cupiennius salei* represent more conserved manifestations of Dpp/BMP2/4 signalling than those I have described for *Tetranychus urticae* (Akiyama-Oda and Oda 2006; Dearden and Akam 2001; Jockusch et al. 2000; Jockusch and Ober 2004; Niwa et al. 2000; Prpic 2004a; Prpic et al. 2003; Sanchez-Salazar et al. 1996). However, this conclusion is based on the assumption that patterns observed in both *Cupiennius* and *Tetranychus* faithfully represent Dpp signal transduction by p-Smad-1, but for reasons previously mentioned I consider the *Tetranychus urticae* p-Smad-1 antibody staining to be questionable. Further data with reduced background staining for both species, and wider tetranychid or acarid sampling would be desirable in this regard, adding weight to interpretations - especially of the observed 'anomaly' in *Tetranychus urticae* Dpp deployment relative to other arthropods.

5.4 Cloning *Tetranychus urticae* EGF receptor (TER)

5.4.1 Two putative *Tetranychus* EGFR genes from degenerate PCR

PCR conditions and primer design

I carried out a degenerate PCR screen on genomic *Tetranychus* DNA derived from embryonic and post-embryonic material, using a touchdown PCR program (for the same reasons mentioned in section 5.1.1) with 2.5mM [Mg²⁺], having demonstrated this concentration to be optimal *via* PCR tests at 0.5mM intervals from 1.5mM to 3.5mM [Mg²⁺]. EGFR is a cell surface receptor tyrosine kinase (RTK) subfamily protein, its structure including a short 5' signal peptide whose cleavage initiates transfer to the endoplasmic reticulum (ER), a cysteine-rich extracellular ligand-binding scaffold domain with repeated EGF (CXCXXGF/YXGXXC) motifs, a single-pass trans-membrane domain and an intracellular TK (tyrosine kinase) domain with conserved ATP binding motifs and Thr/Tyr phosphorylation sites (Greenwald 1985; Hsuan et al. 1989; Jorissen et al. 2003; Lepage et al. 1992; Livneh et al. 1985; Lycett et al. 2001; Shilo 2003; Shilo and Raz 1991). I designed 8, 32 or 64-fold degenerate primers targeted against conserved intracellular tyrosine kinase domain coding sequences, specific amino acid motifs (e.g. N'-VKIPVAIK and HRDLAARN-C') identified by aligning multiple EGFR orthologs. The alignment included taxa from among the disparate metazoan phyla for which EGFR orthologs have been identified to date: for example, arthropods *Drosophila melanogaster*, *Anopheles gambiae*, nematodes *Caenorhabditis elegans*, *C. vulgaris*, trematode *Schistosoma mansoni*, bivalve mollusc *Mytilus*, and vertebrates including chick *Gallus gallus*, fish *Takifugu rubripes* and mammals *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* (Gellner and Brenner 1999; Greenwald 1985; Lax et al. 1988; Livneh et al. 1985; Lycett et al. 2001; Shoemaker et al. 1992; Yamamoto et al. 1984; Yamamoto et al. 1983). See Appendix 2.6 for complete primer sequences.

Target and non-target gene amplification: TER-a, TER-b and Tu-Src

Two of six possible primer pairings gave PCR amplification products at the predicted size of 300-350bp. By analysing multiple clone sequences with BLAST, I identified partial coding sequence for two apparently distinct *Tetranychus* EGFR genes (Figure 5.4.1a). The first, provisionally named 'TER-a' (*Tetranychus* EGF receptor-a) was amplified using degenerate primers EGFR-F1 (5'-AAG AAC GTG AAG ATC CCC GTN GCN ATH AA-3') and EGFR-R2 (5'-TCC GGG CGG CCA RRT CNC KRT G-3'), whereas version 'TER-b' was amplified with EGFR-F2 (5'-AAC GTG AAG ATC CCN GTN GCN AT-3') and EGFR-R1 (5'-CGT GCT TGG GTG TCT GCA CAN RNA CRT T-

(i) *Tetranychus EGF Receptor-a (TER-a)*

```

1/1                               31/11
AAG AAC GTG AAG ATC CCC GTG GCT ATC AAA ATT TTG AAA GAA GGC ACT GCT CCA GAT ACA
lys asn val lys ile pro val ala ile lys ile leu lys glu gly thr ala pro asp thr
61/21                               91/31
AAC AAA GAG TTT CTT GAA GAG GCT TAC ATC ATG GCA TCA GTC GAT CAT CCT AAT TTA CTT
asn lys glu phe leu glu glu ala tyr ile met ala ser val asp his pro asn leu leu
121/41                               151/51
AAG CTT TTA GCT GTT TGT ATG ACT TCA CAA CTC ATG TTG GTT ACT CAA CTT ATG CCT TTA
lys leu leu ala val cys met thr ser gln leu met leu val thr gln leu met pro leu
181/61                               211/71
GGG TGT CTT TTA GAT TAT GTG AAA ACA CAC AAA GAT AAA ATT GGC TCA AAA CCT TTA TTA
gly cys leu leu asp tyr val lys thr his lys asp lys ile gly ser lys pro leu leu
241/81                               271/91
AAT TGG TGC ACT CAA ATT GCT AGA GGA ATG GCT TAT CTT GAA GAT AAA CGA ATG GTT CAC
asn trp cys thr gln ile ala arg gly met ala tyr leu glu asp lys arg met val his
301/101
CGC GAC CTG GCC GCC CGG
arg asp leu ala ala arg

```

(ii) *Tetranychus EGF Receptor-b (TER-b)*

```

1/1                               31/11
AAC GTG AAG ATC CCG GTG GCC ATC AAG GTC CTC CGA GAA GGG ACG CAG CCG AAT ACG AAC
asn val lys ile pro val ala ile lys val leu arg glu gly thr gln pro asn thr asn
61/21                               91/31
AAG GAA TTC CTT GAG GAA GCC TAC ATC ATG GCC AGT GTG GAT CAT CCG AAC GTG CTC AAG
lys glu phe leu glu glu ala tyr ile met ala ser val asp his pro asn val leu lys
121/41                               151/51
CTT TTG GCC GTC TGC ATG ACC TCC CAA CTG ATG CTT GTC ACC CAG CTG ATG CCG CTG GGT
leu leu ala val cys met thr ser gln leu met leu val thr gln leu met pro leu gly
181/61                               211/71
TGT CTC TTG GAC TAC GTG CGA AAC AAC AAA GAC AAA ATC GGT TCG AAA CCT CTA CTC AAC
cys leu leu asp tyr val arg asn asn lys asp lys ile gly ser lys pro leu leu asn
241/81                               271/91
TGG TGC ACC CAA ATC GCT CGG GGC ATG ATG TAT CTA GAG GAG AAA CGC ATG GTG CAC AGA
trp cys thr gln ile ala arg gly met met tyr leu glu glu lys arg met val his arg
301/101                               331/111
GAT CTC GCC CTC CGT AAC GTT CTG GTG CAG ACA CCC AAG CAC
asp leu ala leu arg asn val leu val gln thr pro lys his

```

(iii)

TER-a	ILKEGTAPDTNKEFLEEAYIMASVDHPNLLKLLAVCMTSQMLMLVTQLMP
TER-b	VLREGTQPNTNKEFLEEAYIMASVDHPNVLLKLLAVCMTSQMLMLVTQLMP
TER-a	LGCLLDYVKTHKDKIGSKPLLNWCTQIARGMAYLEDKRMV
TER-b	LGCLLDYVRNNKDKIGSKPLLNWCTQIARGMMYLEEKRMV

Figure 5.4.1a Nucleotide and amino acid sequences for putative EGFR orthologs obtained by degenerate PCR. (i) *TER-a*, primers F1 and R2 in grey. (ii) *TER-b*, F2 and R1 primers in grey. (iii) TER-a and TER-b amino acid code comparison. Yellow: 10/89 differences = 11.2%, thus giving 89.8% protein identity for this short sequence.

3'). Comparison of nucleotide and amino acid level data for the two putative *Tetranychus EGFR* genes (see Figure 5.4.1a) reveals sequence identity differences tabulated below:

Table 5.4.1a Sequence differences between putative *TER-a* and *TER-b* genes

CHARACTERS COMPARED <i>TER-A</i> VS. <i>TER-B</i>	# RELATIVE CHANGES / TOTAL # CHARACTERS	% SEQUENCE DIFFERENCE
Nucleotides	77/267	28.8
Nts: 1 st position	15/267	5.6
Nts: 2 nd position	6/267	2.9
Nts: 3 rd position	56/267	20.9
Amino acids	10/89	11.2

During the *EGFR* screening process, degeneracy of primers caused amplification of one non-target gene; a putative *Tetranychus Src* ortholog. *Src* genes were named after Schmidt-Ruppin viral oncogene/avian sarcoma virus and the *Src* gene family comprises several related oncogenes, some of which - including *EGFR*, *Src* and *abelson* - display receptor tyrosine kinase (RTK) activity (Hoffmann et al. 1983; Livneh et al. 1985; Yamamoto et al. 1983). See Appendix 3.1 for details of *Tu-Src* sequence affinity and analysis.

Multiple sequence analysis of Tu RTK orthologs

I constructed a multiple sequence alignment to compare 90 amino acids from consecutive regions of orthology within the intracellular tyrosine kinase domain of EGFR and EGFR-related receptor tyrosine kinases Src and Abelson (Figure 5.4.1b). I included Src and Abelson to confirm or refute putative *Tu-Src* identity (c.f. Appendix 3.1) and to serve as outgroups for the putative EGFR orthologs compared (Hoffmann et al. 1983; Livneh et al. 1985). Within the EGFR orthology group, 32% (29/90) of residues are identical, whereas shared identity is reduced to 22% (20/90) when all EGFR, Src and Abelson orthologs are compared. An initial visual analysis based on these shared identities supports TER-a + TER-b affinity to other EGF receptors (Figure 5.4.1b).

5.4.2 Extension of *TER-a* sequence data by cDNA library screening and inverse PCR methods

i) *TER-a* cDNA library screening results

I recovered 624bp *TER-a* coding sequence by screening the *Tetranychus urticae* embryonic cDNA library with probe synthesised from the original ~350bp *TER-a* fragment¹⁸, re-amplified from plasmid DNA using degenerate primers EGFR-F2 + EGFR-R1. Novel *TER-a* sequence data obtained by screening includes 345bp coding for 115 amino acids downstream of the previously characterised probe sequence (see Figure 5.4.2a for illustration of screening-derived sequence within the total *TER-a* data retrieved).

ii) *TER-a* inverse PCR results

I recovered 1461bp of un-interrupted *TER-a* open reading frame by inverse PCR (see Chapter VIII, section 8.2.3 for methodology), representing part of the intracellular tyrosine kinase domain (Figure 5.4.2a). The sequence codes for conserved putative ATP binding motifs, and comparison with *Drosophila*, *Anopheles* and human EGFR orthologs suggests that this portion of *TER-a* ORF runs from approx. 50 amino acids C-terminal of the trans-membrane domain to approx. 100 amino acids before a conserved Tyr residue for kinase phosphorylation (Cooke et al. 1987; Hsuan et al. 1989; Livneh et al. 1985; Lycett et al. 2001). Outermost inverse PCR primers used to amplify outwards across novel circularised *TER-a* DNA were iTERb L43 (5' TTG TAT CTG GAC AGT GCC 3') and iTERb U277 (5' TTG AAG ATA AAC GAA TGG T 3'): see Appendix 2.6 for all primer details. Beyond the original, short *TER-a* fragment, 380 novel amino acids (1140bp) were characterised; 235 N-terminal and 145 C-terminal. Of the C-terminal amino acids, 115 matched the sequence length also recovered by cDNA screening (Figure 5.4.2a). In spite of recovering ~1.5Kb CDS by iPCR, transcriptional initiation and termination codons were not encountered, leaving inferences about gene structure incomplete: data for the signal peptide, extracellular, trans-membrane and partial TK domains are missing. CDS for EGFR orthologs in other species are 3.65Kb to 4.3Kb long (e.g. *Drosophila*, 4281bp; *Anopheles*, 4302bp; *C. elegans*, 3972bp; *Mus/Homo*, 3633bp), so we may assume that between 2.15 and 2.8Kb of *Tetranychus EGFR-a* coding sequence may remain further to the 1.5Kb presented here (Cooke et al. 1987; Livneh et al. 1985; Lycett et al. 2001; Schejter 1986; Walker 1998; Yamamoto et al. 1983).

¹⁸ Following excision and sequencing of target inserts in pBK-CMV vector, one clone was identified as anomalous, pertaining not to *EGFR* but to the *Tho Complex (Tho-C)* gene family: see Appendix 3.2 for details.

(i)

1/1 31/11 61/21
ACC AGA GAA GAT GGC GGC TTG AAA TCG TTG TCT GGT AAA TCT GTT TGT CGT AAA TGT CAC CAA AGA TGT AGA AAT TGC ACT GCT TAT GGT
thr arg glu asp gly gly leu lys ser leu lys ser gly lys ser val cys arg lys cys his gln arg cys arg asn cys thr ala tyr gly
91/31 121/41 151/51
ATT CAT AAA TCA GTT TGT GAA TGT TTA CAA TAT TCT TCG GCT GAG CAA TGT GAA GAT ACT TGT CCA CGA GAT CAC TGG GCT GAT ACA GCT
ile his lys ser val cys glu cys leu gln tyr ser ser ala glu gln cys glu asp thr cys pro arg asp his trp ala asp thr ala
181/61 211/71 241/81
AAT CAT GTG TGC TAC AAG TGT GCT GAT GAA TGC CAA GGC TGT CAC GGA CCA ACC AAT GGA GAT TGT TTA TCT TGT CGT AAC TAT CGT GTC
asn his val cys tyr lys cys ala asp glu cys gln gly cys his gly pro thr asn gly asp cys leu ser cys arg asn tyr arg val
271/91 301/101 331/111
TAC TGG GGT GAT GAT CAG AAG CGT TTC AAT TGT ACT GCT ACT TGT CCT GCT GAT AAG CCA TAT AAA GTT GTC GCC GTA AAT GTT GCC GAA
tyr trp gly asp asp gln lys arg phe asn cys thr ala thr cys pro ala asp lys pro tyr lys val ala val asn val ala glu
361/121 391/131 421/141
GAT CCG TAT TGT TCT GAA GAA GAA GCT AAA CCA TTC ATA CCT CCA CCT AAA GAT TCC AAC ACA GCA ATG ATA GTT GGT ATA GCA ATC
asp pro tyr cys ser glu glu glu ala lys pro phe ile pro pro pro lys asp ser asn thr ala met ile val gly val ile ala ile
451/151 481/161 511/171
ACA GTT TTA TGT CTT GGA TTT TTC CTT GCC TTT TAC CAG GGT CTT CAA AAA GCA CGA ACA AAG GAA AAG ACT ATG CAA CTC ACG
thr val leu cys leu gly phe phe leu ala phe phe ser tyr gln gly leu gln lys ala thr lys glu lys thr met gln leu thr
541/181 571/191 601/201
ATG CGA ATG TCC GGG TTC GAA GAT AAT GAG CCT CTT AAA TTG ACT AAT GTT AGG CCA AAC TTG GCT AAA CTC CGC ATT GTA AAA GAG GCA
met arg met ser gly phe glu asp asn glu pro leu lys leu thr asn val arg pro asn leu ala lys leu arg ile val lys glu ala
631/211 661/221 691/231
GAG CTT CGT AAA GGA GGT GTA CTT GGC TGT GGG GCT TTT GGT ATG GTT CAC AAA GGA GTT TGG GTC CCT GAA GGG GAA AAT GTA AAA ATT
glu leu arg lys gly gly val leu glu cys gly ala phe gly met val his lys gly val trp val pro glu gly glu asn val lys ile
721/241 751/251 781/261
CCC GTT GCA ATC AAA ATT TTG AAA GAA GGC ACT GCT CCA GAT ACA AAC AAA GAG TTT CTT GAA GAG GCT TAC ATC ATG GCA AAA GTC GAT
pro val ala ile lys ile leu lys glu gly thr ala pro asp thr asn lys glu phe leu glu glu ala tyr ile met ala lys val asp
811/271 841/281 871/291
CAT CCT AAT TTA CTT AAG CTT TTA GCT GTT TGT ATG ACT TCA CAA CTC ATG TTG GTT ACT CAA CTT ATG CCT TTA GGG TGT CTT TTA GAT
his pro asn leu leu lys leu leu ala val cys met thr ser gln leu met leu val thr gln leu met pro leu gly cys leu leu asp
901/301 931/311 961/321
TAT GTG AAA ACA CAC AAA GAT AAA ATT GGC TCA AAA CCT TTA TTA AAT TGG TGC ACT CAA ATT GCT AGA GGA ATG GCT TAT CTT GAA GAT
tyr val lys thr his lys asp lys ile gly ser lys pro leu leu asn trp cys thr gln ile ala arg gly met ala tyr leu glu asp
991/331 1021/341 1051/351
AAA CGA ATG GTT CAC CGT GAT TTG GCT TTG AGA AAC GTT CTT CTT CAA ACT CCA GGA TGT GTA AAA ATT ACA GAT TTT GGT TTA GCA AAA
lys arg met val his arg asp leu ala leu arg asn val leu leu gln thr pro gly cys val lys ile thr asp phe gly leu ala lys
1081/361 1111/371 1141/381
CTA TTA GAT ATT AAT GAA GAC GAA TAT ATC GCT GAA GGA GGC AAA ATG CCT ATT AAA TGG TTA GCT CTT GAA TGT ATC CAC CAT CGA ATA
leu leu asp ile asn glu asp glu tyr ile ala glu gly lys met pro ile lys trp leu ala leu glu cys ile his his arg ile

Figure 5.4.2a - Page 1 of 2

1171/391 1201/401 1231/411

TTT ACT CAT AAA AGC GAT GTT TGG GCT TTT GGT ATT ACT GTT TGG GAA CTT TTA ACA TAT GGC GGA CGT CCA TAT GAA GGT GTA AAT GCT
phe thr his lys ser asp val trp ala phe gly ile thr val trp glu leu thr tyr gly gly arg pro tyr glu gly val asn ala
1261/421 1291/431 1321/441

AGA GAA GTT CCA TCC CTT TTA GAT AAA GGT GAG CGG CTT CCA CAG CCC GCA ATT TGT ACA ATA GAT GTT TAC ATG ATC ATG ATT AAG TGT
arg glu val pro ser leu leu asp lys gly glu arg leu pro gln pro ala ile cys thr ile asp val tyr met ile met ile lys cys
1351/451 1381/461 1411/471

TGG ATG CTG GAT GCA GAG TCT CGT CCA TCC TTT AGA GAC TTG GCT ATG GTG TTT GCC AAA ATG GCT GGT GAT CCT GGA AGA TAT CTC GTC
trp met leu asp ala glu ser arg pro ser phe arg asp leu ala met val phe ala lys met ala gly asp pro gly arg tyr leu val
1441/481

ATT CCA GGA GAC AAG TTT ATG
ile pro gly asp lys phe met

(ii)

1 11 21 31 41 51 61 71
TREDGGLKSL SGKSVCRKCH QRCRNCTAYG IHKSVCECLQ YSSAEQCEDT CPRDHWADTA NHVCYKCADE CQCHGPTNG
81 91 101 111 121 131 141 151
DCLSCRNYRV YWGDDQKRFN CTATCPADKP YKVAVNVAE DPYCSEEEAK PFIPPPKDSN TAMIVGVIAI TVLCLGFFLA
161 171 181 191 201 211 221 231
FFSYQGLQKA RTKEKTMQLT MRMSGFEDNE PLKLTNVRPN LAKLRIVKEA ELRKGGLVGC GAFGMVHKGV WVPEGENVKI
241 251 261 271 281 291 301 311
PVAIKILKEG TAPDTNKEFL EEAYIMAKVD HPNLLKLLAV CMTSQLMLVT QLMPLGCLLD YVTKHKDKIG SKPLLNWCTQ
321 331 341 351 361 371 381 391
IARGMAYLED KRVVHRDLAL RNVLLOTPGC VKITDFGLAK LLDINEDEYI AEGGKMPIK W LALECIHHRI FTHKSDVWAF
401 411 421 431 441 451 461 471
GITVWELLTY GGRPYEGVNA REVPSLLDKG ERLPQPAICT IDVYMIMIKC WMLDAESRPS FRDLAMVFAK MAGDPGRYLV
481
IPGDKFM

Figure 5.4.2a - Page 2 of 2

(i) *TER-a* 1461bp nucleotide and 481 amino acid sequence obtained by cDNA library screening (**bold sequence**) and inverse PCR (extended sequence). (ii) Summary of *TER-a* amino acid sequence (487 residues). **Boxed**: putative ATP-binding motifs/residues. **Grey**: original short *TER-a* fragment on which iPCR primers are designed; **yellow** = outermost upstream primer L43 (5' TTG TAT CTG GAC AGT GCC 3'); **green** = outermost downstream directed primer U277 (5' TT GAA GAT AAA CGA ATG GT 3').

iii) Multiple sequence alignment and phylogenetic analysis

A multiple alignment comparing 240 homologous amino acids of EGFR and EGFR-related sequences indicates 22.1% (53/240) shared identity between all the proteins (Figure 5.4.2b). I aligned 11 arthropod, vertebrate, nematode and hydroid EGFR orthologs, with 9 arthropod, vertebrate, echinoderm and Src/Abelson proteins as outgroup sequences (Hoffmann et al. 1983; Livneh et al. 1985). *Tetranychus* TER-a has discernable signature EGFR residues, including several motifs specific to arthropod-derived EGF receptors and a number of putative ATP-binding motifs, significant in the ATP phospho-transferase reaction (Figure 5.4.2b) (Hsuan et al. 1989; Krebs and Beavo 1979).

I carried out Bayesian inference analysis of the nucleotide dataset from which the previous protein alignment is derived, producing a robust consensus tree topology with maximal/near-maximal posterior probability support at all nodes (Figure 5.4.2c). The extended *TER-a* sequence is confirmed to be a true *Tetranychus* EGFR ortholog, more closely related to *TER-b* than to other arthropod (dipteran) EGFR genes. *Hydra* TK and *Caenorhabditis* Let25 sequences branch outside a clade of distinct vertebrate and arthropod EGFR genes; their 'basal' placement may be due to genuine gene phylogeny, or may result from an artefact of mutual attraction between the two relatively rapidly evolving, long branch status sequences. This point aside, the topology supports monophyly of EGFR genes, excluding the monophyletic *Src* sister clade and more derived *Abelson* outgroup.

5.4.3 *TER-a* DIG-labelled ssRNA probes

Synthesis of TER-a probes

I synthesised single stranded, DIG-AP labelled mRNA probes from *TER-a* inserts in pGEM®-T Easy or pBK-CMV vector. Inserts in pBK-CMV were excised from the cDNA library, and *TER-a* fragments between 683bp and 949bp in length were cloned into pGEM®-T Easy vector after amplification from *Tetranychus* genomic DNA with gene-specific primers that I designed against targets within the ~1.5Kb *TER-a* coding sequence (Figure 5.4.3a). Insert size was confirmed by (restriction enzyme) RE digesting 1µl purified plasmid with EcoRI (single digest) or NotI/EcoRI (double digest) for inserts in the respective pGEM®-T Easy and pBK-CMV vectors, avoiding RE cut sites within the gene itself (for which see Figure 5.4.3a,ii). For ssRNA probe synthesis, I selected specific restriction enzymes for plasmid linearization and specific RNA polymerases for mRNA transcription, as appropriate to (i) the particular vector plasmid's multiple cloning site, (ii) *TER-a* insert orientation in the vector and (iii) which RNA probe polarity (sense/anti-sense) was required from the transcription synthesis reaction (Figure 5.4.3b).

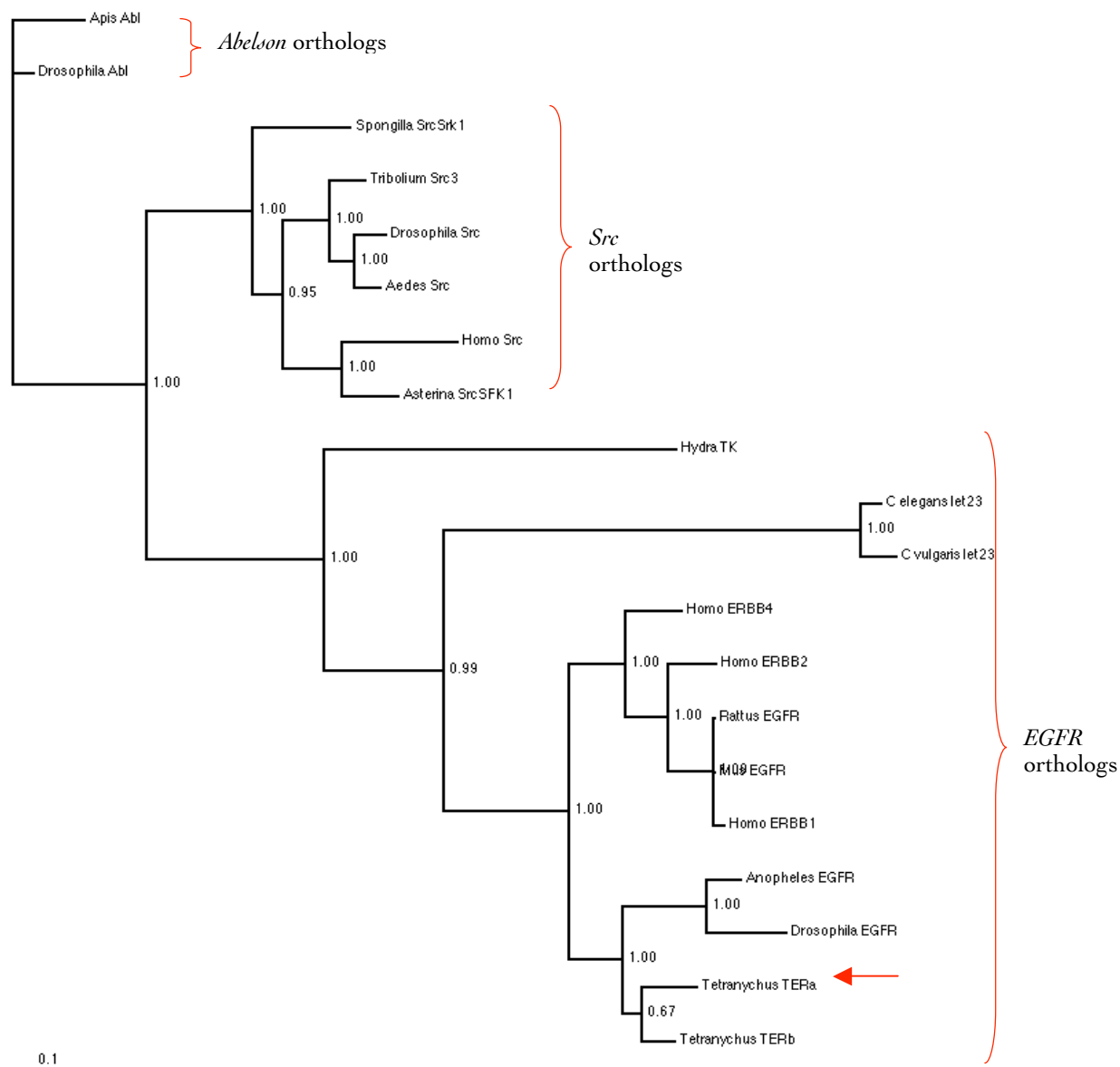


Figure 5.4.2c Bayesian inference topology from analysis of 720bp *Tu-EGFR-a* coding sequence, with *EGFR*, *src* and *abelson* orthologs from a range of other taxa. *TER-a* (red arrow) clearly groups with other arthropods in a fully supported *EGFR* orthology clade, and the short *TER-b* sequence branches as a duplicate gene with distinct branch length properties. Both outgroup clades receive maximal posterior probability support and indicate closer relatedness of *Src* genes to *EGFR* than *Abelson* genes to *EGFR* (which may explain amplification of *Tu-Src* but no *Tu-Abl* during degenerate PCR screening: c.f. section 5.4.1). Scale bar: 0.1 changes expected per base pair position in sequence alignment.

(i) (1) ACCAGAGAAGATGGCGGC**TTGAAATCGTTGTCTGGT**AAATCTGTTTGTCGTAAATGT
 CACCAAAGATGTAGAAAT**TGCAC**TGCTTATGGTATT**CATAAA**TCAGTTTGTGAATGTTT
 ACAATATTCTTCGGCTGAGCAATGTGAAGATAC**TTGTCC**CACGAGATCACTGGGCTGATA
 CAGCTAATCATGTGTGCTACAAGTGTGCTGATGAATGC**CAAGGCTGT**CACGGACCAACC
 AATGGAGATTGTTTATCTTGTCTGTAACATCGTGTCTACTGGGGTGATGATCAGAAGCG
 TTTCAATTGTACTGCTACTTGTCTCTGCTGATAAGCCATATAAAAGTTGT**CGCCG**TAAATG
 TTGCCGAAGATCCGTATTGTTCTGAAGAAGAAGCTAAACCATTTCATACCTCCACCTAAA
 GATTCCAACACAGCAATGATAGTTGGTGTTATAGCAATCACAGTTT**TATGCT**TGGATT
 TTTCTTGCCTTT**TTCAGTTACCAGGGTCTT**CAAAAAGCACGAACAAAGGAAAAGACTA
 TGCAACTCACGATGCGAATGTCCGGGTT**CGAAGATAAT**GAGCCTCTTAAATTGACTAAT
 GTTAGGCCAAACTTGGCTAAACTCCGCATTGTAAAAGAGGCAGAGCTTCGTAAAGGAGG
 TGTACTTGGCTGTGGGGCTTTTGGTATGGTTCACAAAGGAGTTGGGGTCCCTGAAGGGG
 AAAATGTAAAAATTCCCGTTGCAATCAAAATTTGAAAGAA**GGCAC**TGCTCCAGATACA
 AACAAAGAGTTTCTTGAAGAGGCTTACATCATGGCAAAAGTCGATCATCCTAAATTTACT
 TAAGCTTTTAGCTGTTTGTATGACTTCACAACCATGTTGGTTACTCAACTTATGCCTT
 TAGGGTGTCTTTTAGATTATGTGAAAACACACAAGATAAAATTTGGCTCAAAACCTTTA
 TTAAATTGGTGCACCTCAAATTGCTA**CACCAATGGCTTATCTTC**AAGATAAACGAATGGT
 TCACCGTGATTTGGCTTTGAGAAACGTTCTTCTTCAAAC**TCCAGGATGTGT**AAAAATTA
 CAGATTTTGGTTTAGCAAACTATTAGATATTAAATGAAGACGAATATATCGCTGAAGGA
 GGCAAAATGCCTATTAAATGGTTAGCTCTTGAATGTATCCACCATCGAATATTTACTCA
 TAAAAGCGATGTTTGGGCTTTTGGTATTACTGTTTGGGAACTTTAAACATATGGCGGAC
 GTCCATATGAAGGTGTAATGCTAGAGAAGTTCATCCCTTTTAGATAAAGGTGAGCGG
 CTTCCACAGCCCGCAATTTGTACAATAGATGTTTACATGATCATGATTAAGTGTGGAT
 GCTGGATGCAGAGTCTCGTCCATCCTTTAGAGACTTGGCTATGGTGT**TGGCCAAATGG**
 CTGGTGATCCTGGAAG**ATATCTCTCATTCACG**AGACAAGTTTATG (1461)

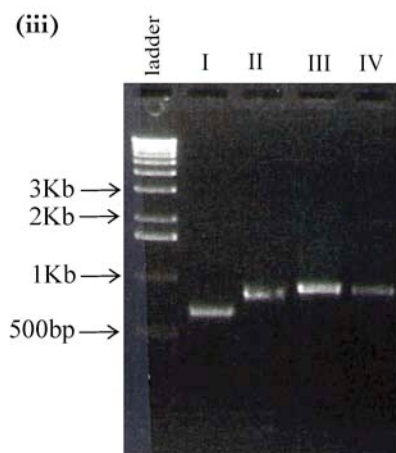
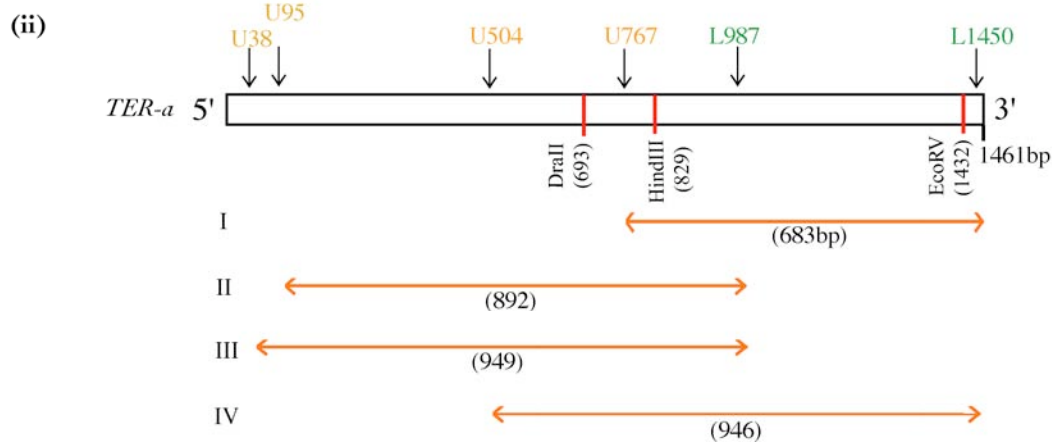


Figure 5.4.3a

(i) Available *TER-a* coding nucleotide sequence.

Highlighting: forward (yellow) and reverse (green) primers, used to amplify ISH probe material from genomic DNA.

(ii) Scaled schematic representation of *TER-a* gene sequence; incomplete CDS for the intracellular RTK domain. Orange/green arrows refer to primers shown in (i), with fragment sizes for associated PCR products I - IV.

(iii) Agarose gel run to show *TER-a* products that correspond to DNA fragments I- IV in (ii).

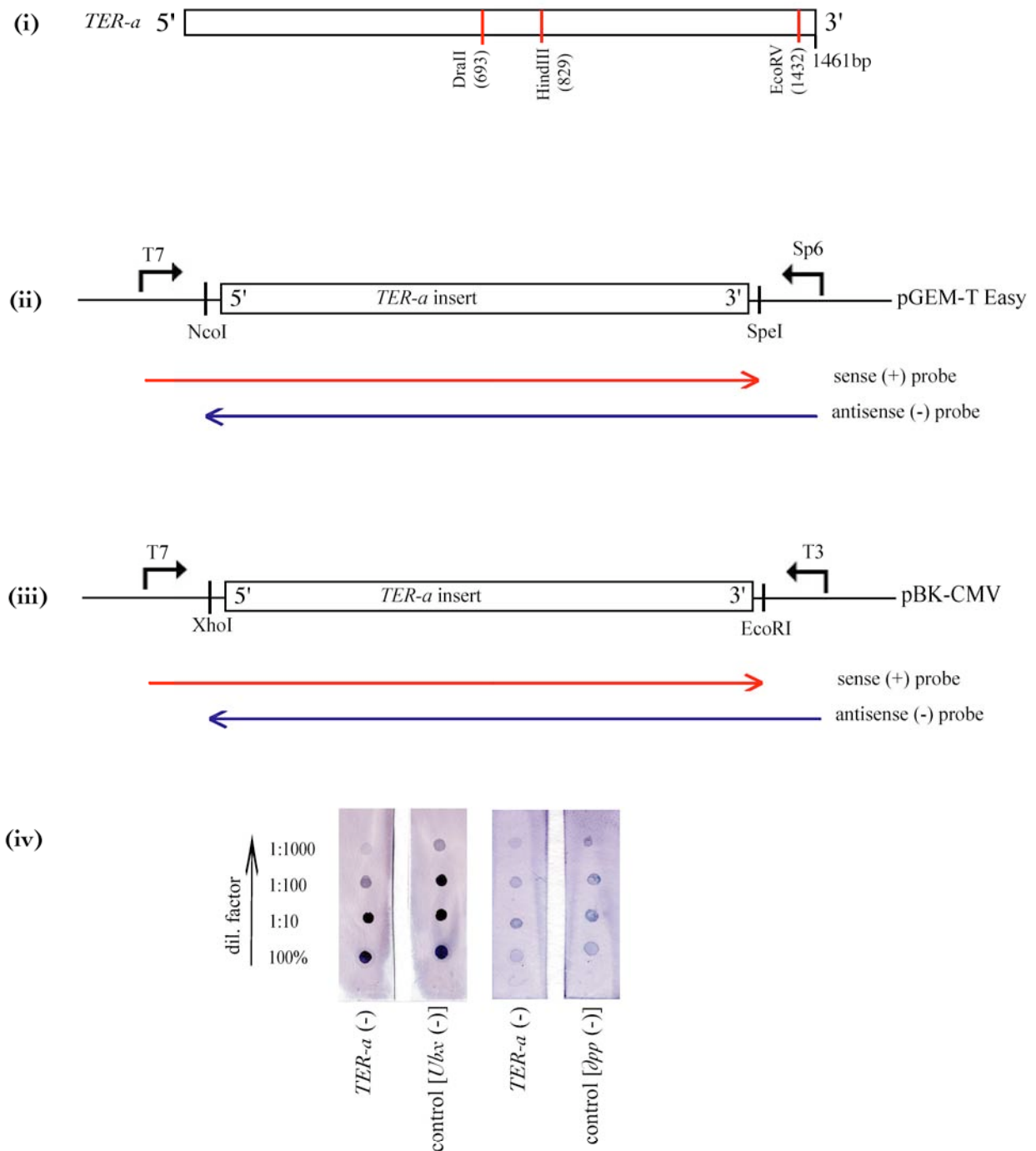


Figure 5.4.3b *TER-a* ssRNA probe synthesis:

(i) Scaled schematic of *TER-a* CDS obtained, with restriction enzyme (RE) cut sites avoided in plasmid linearisation and RNA transcription.

(ii)-(iii) Schematic diagrams of *TER-a* inserts in (ii) pGEM-T Easy and (ii) pBK-CMV vector, with REs and RNA polymerases appropriate for synthesis of sense (red arrow) and antisense (blue arrow) RNA probe. The *TER-a* insert is shown in 5'→3' orientation; reverse reagent combinations for inserts in 3'→5' direction.

(iv) Dot blot hybridisation tests to show effective DIG-labelling of ssRNA probes: decreasing signal with increasing probe dilution = positive result.

Verifying digoxigenin labelling

Effective α -DIG-AP labelling of probe RNA was verified as in previous instances, by testing serially diluted probe samples for phosphatase activity (c.f. Chapter VIII: Materials and Methods, section 8.4.2). The results for these assays were positive, confirming decreased phosphatase reaction with decreased concentration of labelled probe – due to reduced substrate for the correlated pigment synthesis reaction (Figure 5.4.3b,iv).

5.4.4 Attempts to extend *TER-b* sequence by cDNA library screening and iPCR

i) *TER-b* cDNA library screening: negative results

TER-b DNA for use in probe synthesis was amplified from ~350bp inserts carried in purified pGEM®-T Easy vector plasmids, using EGFR-F2 and EGFR-R1 primers that had amplified *TER-b* in initial degenerate PCR on genomic DNA. Following primary, secondary and tertiary screening with only one of the three HRP-labelled *TER-b* probes tested, an appropriate number of luminescent plaques were observed and selected (the other two showed no positive plaques at primary screening stage). PCR on phage picks with primers EGFR-F1 and EGFR-R1 confirmed the presence of an appropriately sized insert (Figure 5.4.4a,i), assumed to indicate target *TER-b*. However, upon excision and sequencing it became clear that the probe had bound to non-target phage DNA (Figure 5.4.4b): vector sequencing primers (T7/T3) amplified almost complete CDS for RNA polymerase II Ssu 72 (119 N' fragment + 79 C' fragment = 198 amino acids, similar to *Mus* Ssu 72: 182, *Homo*: 194 and *Saccharomyces*: 206 amino acids), and *TER-b* specific primers, designed to amplify from known sites within the gene (e.g. L60: 5' CTA GAG GAG AAA CGC ATG G 3' and U285: 5' ATT CCT TGT TCG TAT TCG G 3'), did not work (Altschul et al. 1990).

ii) *TER-b* inverse PCR: negative results

Sequence data further to that obtained for *TER-b* during degenerate PCR screening was not obtained by inverse PCR, in spite of testing 24 primer combinations, different PCR conditions and multiple circularised DNA templates. Genomic extracts of embryonic and adult/nymph-derived DNA were digested separately by four restriction enzymes (BamHI, EcoRV, NcoI and XhoI), giving 8 distinct circularised gDNA template sources. Nested PCR conditions included either i) 56-50°C touchdown + 50°C annealing, ii) 54-49°C touchdown + 49°C annealing or iii) 54-48°C touchdown + 48°C annealing temperature, with iv) 3.0mM or 2.5mM $[Mg^{2+}]$ reagent mix concentration. Under the

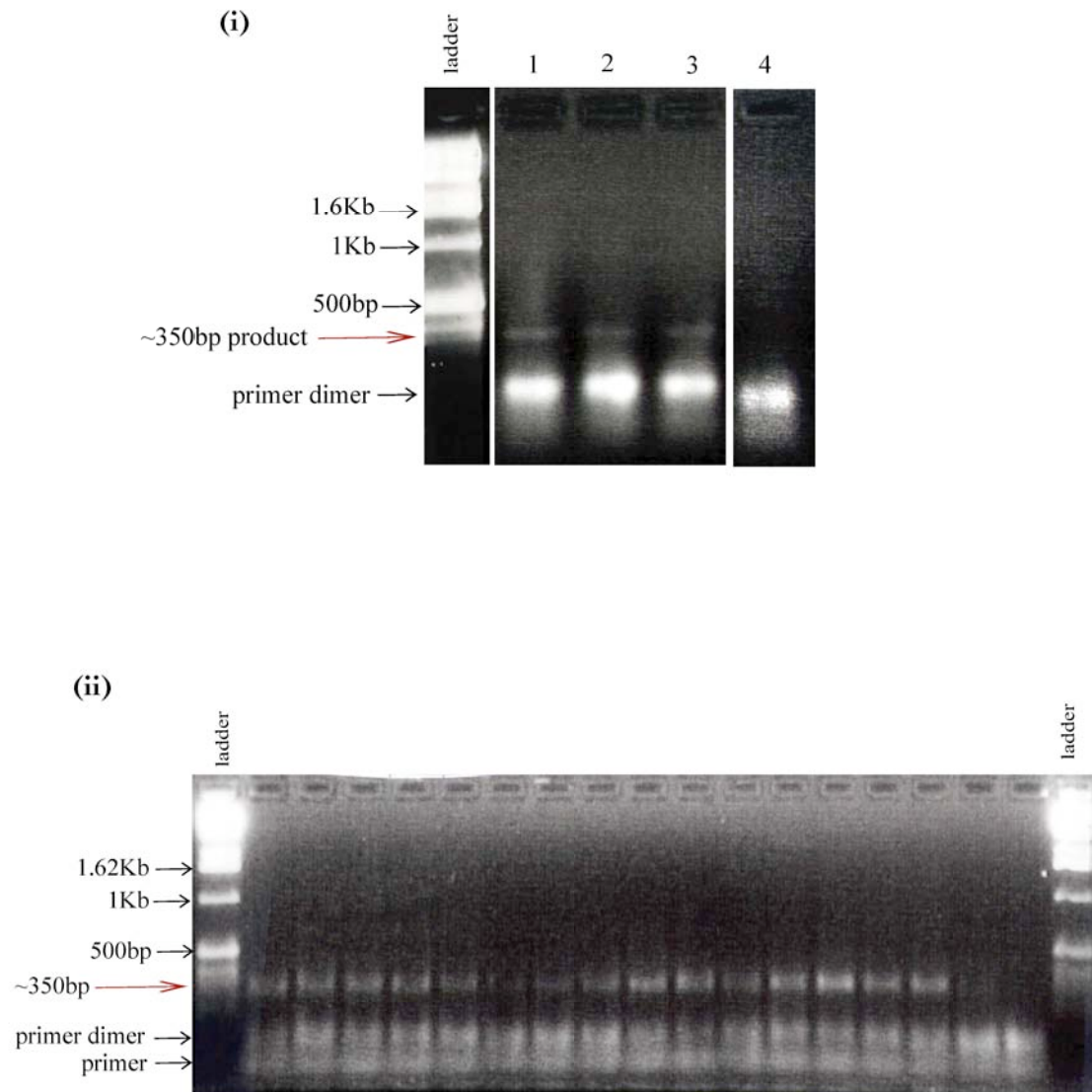


Figure 5.4.4a (i) Primary screening pick PCR with EGFR primers: 1% agarose gel run of DNA products. Lanes 1 - 3: putative 350bp EGFR product (red arrow); Lane 4: negative control. **(ii)** Tertiary screening pick PCR, putative *TER-a* DNA fragments run at expected size for primers used (~350bp). Phage picked from among abundant positive plaques exposed by chemiluminescence on nitrocellulose membranes, but sequencing excised phage yielded non-*TER-b* results. Gel bands in (i) and (ii) may thus be non-target DNA, recognised and perpetuated from primary screening onwards.

(i)

```
1/1      31/11      61/21
GCA CGA GGG AGA TCA GGT GTT ATT TAA TTA CAG ATT CAA TTG GTT AAA TTG TTT TAA TTT ATT TGT AAT TAT TCA TTA ATA CTT TCC TAA

91/31    121/41    151/51
TCT TTG TAT TAT CTG TAC AGT GTA GTG TTG ATT AGT TCT TAA CTT GAT CCA TTC GAT TAC TAA TTG ATT CTC TCT TCC TTT

181/61   211/71
AAA AAA ATG ACT GGT AGA GAT TTA AGA CTA GAT GTT GTT TGC TCT AAC CAA AAT CGA AGT ATG GAA GCT CAT CAA GTC CTT AAT AAG
    met thr gly arg asp leu arg leu ala val val cys ser ser asn gin asn arg ser met giu ala his gln val leu asn lys
271/91   301/101
AAA GGT TTC AAT GTA CGA TCT TTT GGA TCA GGT AAT TGT GTA AAA TTA CCT GGA CCT TCA TTT GAC AAA CCA AAT ATA TAT GAT TTT AAT
lys gly phe asn val arg ser phe gly ser gly asn cys val lys leu pro gly pro ser phe asp lys pro asn ile tyr asp phe asn
361/121   391/131
ACT ACA TAT GAT CAA ATG TAT ACT GAT TTG ATA AAC AAA GAT AAG ACT TTA TAC ACT CAA AAT GGG ATT CTT CAC ATG TTG GAC CGC AAT
thr thr tyr asp gln met tyr thr asp leu ile asn lys asp lys thr leu tyr thr gln asn gly ile leu his met leu asp arg asn
451/151   481/161
AGG AGA ATT AAA CAA AAG CCA GAG AGA TTT CAC GAC TGT TTT GAT AAA TTT GAT GTC ATA TTT ACT TGT GAA GAA AGA GTC TAT GAT CAA
arg arg ile lys gln lys pro giu arg phe his asp cys phe asp lys phe thr cys glu arg val tyr asp gln

541/181   GTA ATT
val ile
```

(ii)

```
1/1      31/11      61/21
GGC GTT GAA GAT CGA TAT TTC TCT GTA GTT CAC CCT GTA AAT GAT AAT GGT GTT GAT GGT ACA AAT GGC AAT GGC AAT CAA GGT
gly val giu asp arg tyr phe ser val val his pro val asn asp asn gly val asp gly thr asn gly thr asn gln gly

91/31    121/41
AAT CAA GGT CCA ATT AAT TCG GTT GCA GTT GAT GAT GGT GAA TCA AAT GAG CAA AAT GAT GAC CAC ACC AAT CAA AGT TTG GAT
asn gln gly pro ile asn ser val ala val asp asp gly asn gly glu ser asn glu gln asn asp asp his thr asn gln ser leu asp
181/61   211/71
AAA GAT TTA ACA GGC TCA AAT AAT GGA GCA AAT GGA TCA CCA GGA TCT GGC TCT TAA CAA ATC TTC AAA TTT TTG GTT CCT CTC ATT
lys asp leu thr gly ser asn asn gly ala asn gly ser pro gly ser -O-

271/91   301/101
TTT ATC TGC TCT CTT TTT TTT TAC CAA TTA CAG AAG TTA ATT CTA ATT GTC GAT GAA CCA ACT TAA ACA ATC TTG CTC AAA TCA AGC

361/121   391/131
AAT TAT CTT TCT TCT TCT CTC TGG CAT CAT TAT CAT TTT CGT TTA ATC TAT TGG TAG TTT ACA ATC GTT TTT AGC TGT CTA GAG

451/151   481/161
AAG GTT AAT TGT TAA TAT ATA ATA CAT ATT ATT ATT AAT ACC TAT TTG ATG GAT TTT ACT GTA AAT CGT ATT TCC AAT TAT TCA AAA AAA

541/181   AAA AAA AAA
```

Figure 5.4.4b Examples of products obtained when cDNA library screening with putative *TER-b* probe. (i) Small sub-unit 72 (Ssu72) RNA polymerase II: 5' UTR and 5' CDS. (ii) Sequence unidentified by BLAST but potentially 3' CDS for Ssu72 RNA pol-II, as (i) and (ii) were sequenced from either side of an insert into one pBK-CMV plasmid clone. **met** – start codon, **O** – stop codon.

conditions: 50°C annealing + 3.0mM [Mg²⁺], the primer combinations L143/U160 for primary iPCR + L101/U160 for nested iPCR amplified several potentially positive iPCR products between 500bp to ~1.2Kb (Figure 5.4.4c). In three cases multiple fragments were amplified within the same reaction, requiring separation by slow gel electrophoresis prior to purification (Figure 5.4.4c,ii). Sequencing results (Figure 5.4.4d) yielded no *EGFR*-related genetic information: BLAST analysis (Altschul et al. 1990) indicated that all the potential *TER-b* iPCR products I cloned were artefacts pertaining to:

- i) a plant protein (related to *Phaseolus*, *Arabidopsis* or *Medicago* DNA polymerase, GAG proteins, RT kinase and peptidase),
- ii) a *Burkholderia* bacterial membrane protein, and
- iii) various cloning vector regions.

Plant and bacterial affinities may be explained by the presence of contaminant material in the original genomic DNA extract, derived from unavoidable bean leaf debris and digestive, symbiotic bacteria within the spider mites. Successful inverse PCR amplification of *TER-a* coding sequence from the same DNA extracts (section 5.4.2) indicates a sufficient and significant proportion of intact *Tu*-DNA within the templates, however, regardless of any plant/bacterial contamination. This observation indicates that, contrary to poor quality or contaminated DNA being implicated, negative *TER-b* results were obtained due to problems with primer binding (specificity or efficiency), or erroneous assumptions about the initial presumed *TER-b* sequence itself.

5.4.5 Critical assessment of *TER-b* authenticity

Summary of putative TER-b results

Using degenerate primers EGFR-F2 and EGFR-R1, I amplified a 342bp gene fragment from genomic DNA template (Figure 5.4.1a). BLAST analysis, sequence alignment and Bayesian phylogenetic analysis confirmed this as an arthropod *EGFR* ortholog, logically assumed to be part of the *Tetranychus* genome and named *TER-b* to indicate a second putative *EGFR* paralog (Figure 5.4.1b) (Altschul et al. 1990). However, as described in section 5.4.4 neither cDNA library screening nor inverse PCR on *Tetranychus* genomic DNA provided any positive results or further sequence data.

Assessment of negative results

Regarding the *Tetranychus* cDNA library screen, I tested probes from three different clones, one binding non-specifically to a *Ssu72 RNA polymeraseII* gene, and others evidently failing to bind to any phage DNA even during primary screening. This indicates that the *TER-b* gene was unrepresented in the cDNA library, either due to lambda phage degradation, or because the original *TER-b* fragment

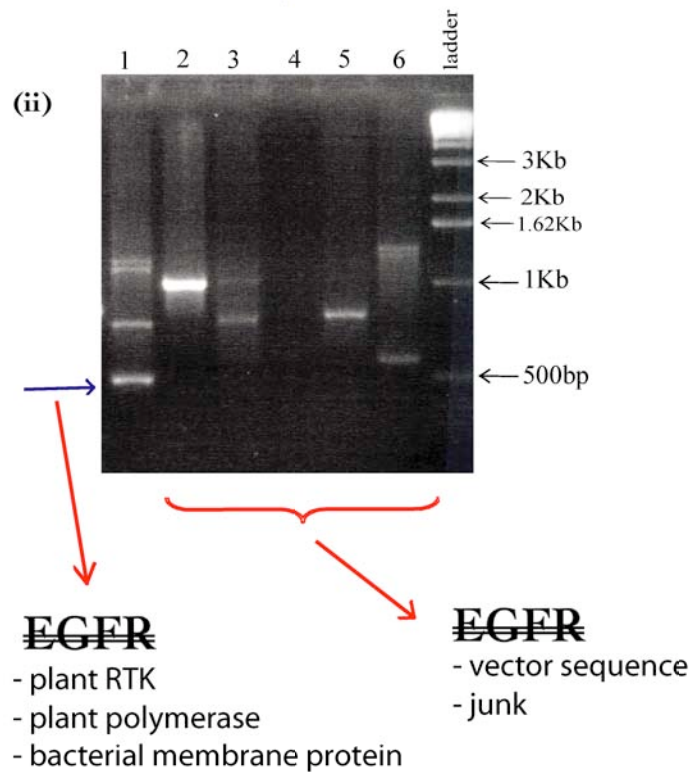
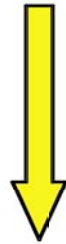
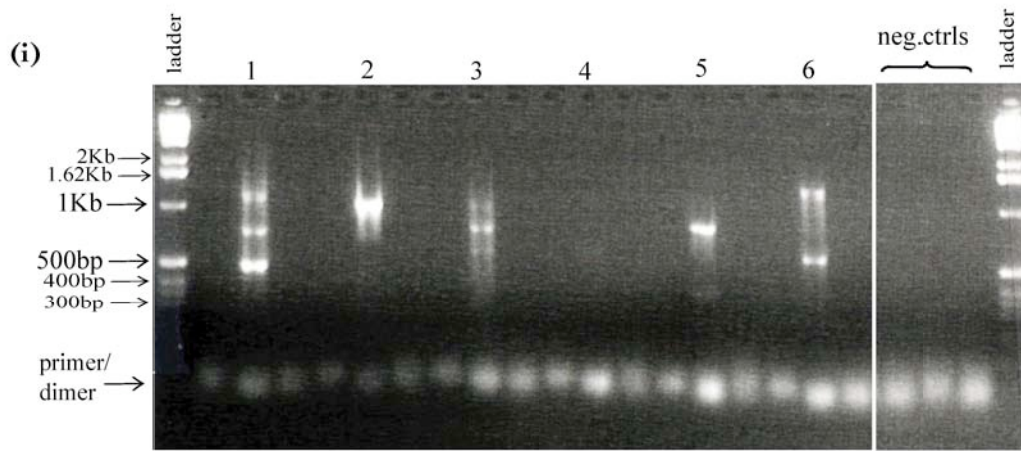


Figure 5.4.4c

(i) Agarose gel run of final *TER-b* inverse PCR reactions; single or multiple putative products only obtained for 6 specific primer combinations - lanes marked 1 - 6 for cross-reference with (ii). iPrimer details in text.

(ii) Slow run agarose gel of lane 1 - 6 DNA products, permitting separate extraction and cloning of individual DNA fragments from among multiple bands. Sequencing (c.f. Figure 5.4.4d) showed all products to be non-*Tetranychus* genes or vector material.

(i)

```
1/1 1/1 31/11 61/21
ACA ACC CAG CGG CAT CAG GAC GCG TTT GCT TGG TCC GCT TCG GAC ATG CCC GGC ATC GAT CCC GAC TTC TTG TGC CAT CAT CTG GCG ATG
thr thr gln arg his gln asp ala phe ala thr ser ala ser asp met pro gly ile asp pro asp phe leu cys his his leu ala met
91/31 121/41
GAC AAC ATG GTG AGA CCG GTG CGA CAA AGA AGA AAA TTC AAT GAG GAG AGG AGG CAG CGC ATT AGG GAC GAA ACA CAA AAA CTC CTC
asp asn met val arg pro val arg gln arg arg lys phe asn glu glu arg arg gln ala ile arg asp glu thr gln lys leu leu
181/61 211/71
GCT GCA GGC CAC ATC AGG GAA GTC CAG TAC CCT GAA TGG TTG GCA AAT GTC GTA CTG GTG AAG AAG AGT AAC GGG AAA TGG CGC ATG TGC
ala ala gly his ile arg glu val gln tyr pro glu trp leu ala asn val val leu val lys ser asn gly lys trp arg met cys
271/91 301/101
GTC GAT TTC ACC GAT CTG AAC GCT TGC CCA AAG GAT TCG TAT CCT TTA CCG AGC ATT GAT GCC CTG GTT GAT AGT GCT GCA GGG TGC
val asp phe thr asp leu asn lys ala cys pro lys asp ser tyr pro leu pro ser ile asp ala leu val asp ser ala ala gly cys
361/121 391/131
AAG CTG CTG AGT TTC CTG GAT GCC TTC TCG GGC TAT AAT CAG ATC AAG ATG CAT CCC ATG GAT GAA GAA AAA ACA GCC TTC ATG ACG GAG
lys leu leu ser phe leu asp ala phe ser gly tyr asn gln ile lys met his pro met asp glu glu lys thr ala phe met thr glu
451/151 481/161
AGG TCA TGC TAC TGC TAT AAG GTG ATG CCG CTG GGT TGT
arg ser cys tyr cys tyr lys val met pro leu gly cys
```

(ii)

```
1/1 31/11 61/21
GAC AAC CCA CCG GCA TCA CCG CGC CTG TCC GTC AGT CCG TCG CCT ATA CTC AAA TCA CGC GGC CGC TTC CAT CAC AAC TAC AGC GAG CCT
asp asn pro ala ala ser arg arg leu ser val ser pro pro ile leu lys ser arg gly arg phe his his asn tyr ser glu pro
91/31 121/41
TGC ACA GCA CCG TAT CAG GAG ACA AAT ATG ACC ACC TAC TTC ACG ATC GGT GAT TTC ATT CTG GTC ATT CCG ATG GCG CTA GCC GGC GCG
cys thr ala ala tyr gln glu thr asn met thr thr phe thr phe thr gly asp phe ile leu val ile pro met ala leu ala gly ala
181/61 211/71
CTC TTT CTT GGC GCG TGC CCG TGT GCA ACC GAG TTC AGG CAT AAC GTG CTG CCG GTA TTG GGC GCA TTG CTC GGT GTG GTC GTC GCG GTG
leu phe leu gly ala cys pro cys ala thr glu phe arg his asn val leu arg val leu gly ala leu leu gly val val ala val
271/91 301/101
CTG CTG GTC GAA GGA CTA CCC GTA CTG GTC TAG CGA CCC GAC TAC TCA AAT CAC CGG AGC AGG CGC TCC GCA TAC GTG ATG CAC ACG TAA
leu leu val glu gly leu pro val leu val -0- 331
361 421
AAA AGC CGC CTT GCG GCG GCT TTC CGG ATA CGC GTG TGC GAT CAG ACA TCA ACC GCA AGG CCC GCT CAG ACG TGC TGA TCG GTC GAC GGC
451
TTT TCC CAC AGG TTG ATG CCG CTG GGT TGT
```

Figure 5.4.4d Sequence data for identifiable genes recovered by cloning hypothetical *TER-b* inverse PCR products. (i) CDS coding for protein with highest BLAST similarity to *Phaeococcus* GAG-polymerase polyprotein. (ii) 3' CDS and UTR encoding protein with strong identity (91%) to *Burkholderia* bacterial membrane protein. No *Tetranichus urticae* *EGFR* sequences were obtained from any of the clones.

pertains to an arthropod other than *Tetranychus urticae* (phylogenetic analysis clearly indicates arthropod origin: c.f. Figure 5.4.2c).

With respect to inverse PCR, I tested different PCR conditions (annealing T°C, [Mg²⁺]), many primer combinations and 4 distinct restriction enzyme digests of 4 separate DNA extractions (i.e. 16 in total). Products were amplified as a result of non-specific binding to vector material in most cases, to a plant *GAG-polymerase* gene in one clone, and in another to a bacterial membrane protein gene. The plant and bacterial sequences (Figure 5.4.4d) have significant affinity to *RTK* genes, suggesting that *TER-b* iPCR primers may have bound to generic *RTK* motifs sufficient to amplify *RTK*-related genes from *Phaseolus* (bean) and spider mite gut bacteria whose DNA was present within the *Tetranychus* genomic DNA samples (Boudreaux 1963; King and Carroll 2001). Lack of *Tetranychus* products indicates either problems with primer binding - which seems improbable considering their specificity and large number tested - or again that the '*TER-b*' sequence on which specific primers were designed was not in fact derived from *Tetranychus* to start with.

Re-examining TER-b experimentally

I carried out touchdown PCR on newly prepared *Tetranychus* genomic DNA and cDNA template, using either the original degenerate primer pair (EGFR-F2+R1) or combinations of six *TER-b* specific primers (sequences in Appendix 2.6). No product was amplified under any of the PCR conditions tested, neither at the predicted size of ~300bp or at any other molecular weight. This negative result supports the conclusion that the proposed *TER-b* sequence derives from non-*Tetranychus* material in the initial genomic preparations from which it was amplified. A possible candidate would be *Phytoseiulus sp.*, a predatory mite known to attack phytophagous spider mites and used in agriculture and horticulture for biological pest control against *Tetranychus urticae* (Bynum and Porter 2006; Collyer 1998; Florida 2005). The extreme similarity in amino acid code between two EGFR orthologs could then be explained by the phylogenetic proximity of the two species *Phytoseiulus* and *Tetranychus*, both being prostigmatan acarids within the Chelicerata. *Phytoseiulus* tends to be present discontinuously within a given prey population, conceivably having been present at the time the initial genomic preparation was made but disappearing later on, allowing more pure *Tetranychus* DNA to be retrieved (Bynum and Porter 2006; Collyer 1998).

5.5 *TER-a* mRNA transcription during *Tetranychus* embryogenesis

5.5.1 Blastoderm to germdisc stage: no *TER-a* transcripts

Lack of TER-a mRNA transcription

I hybridised whole mount spider mite embryos at blastoderm and germdisc stages (9-15hr AEL) with DIG-AP labelled ssRNA probes between ~680bp and 1Kb in length, synthesised from *TER-a* inserts in both pGEM®-T Easy and pBK-CMV vector plasmids. No *TER-a* transcriptional activity was detected at these stages (Figure 5.5.1,i-iv). Minimal interference from background signal/noise, reproducible results and positive hybridisation patterns visualised *via* the same probes at later stages in development, all support this negative interpretation.

Comparison: EGFR in early D/V and CNS patterning

Drosophila EGF receptor (DER) is expressed prior to gastrulation with a role in dorso-ventral CNS patterning to specify intermediate neuroblasts, and once gastrulation commences, in further ventral and anterior neuroectoderm patterning, dependent on signalling from the ventral midline and in association with proneural and BMP gene regulation (Dumstreit et al. 1998; Ip et al. 1992; Mayer and Nusslein-Volhard 1988; Nambu et al. 1991; Rutledge et al. 1992; Shilo 2003; Shilo and Raz 1991; Skeath 1998; Udolph et al. 1998; Yagi et al. 1998). In these contexts EGFR seems to function by promoting proper differentiation, preventing apoptotic cell death and in some cases determining cell fate. Similar functions have been suggested with respect to vertebrate EGFR, in which roles in regulating axial patterning, apoptosis and morphogenesis during neurogenesis and limb development are apparent (De Robertis and Sasai 1996; Dealy et al. 1998; Kim and Crews 1993; Omi et al. 2005; Tickle 2003; Tickle and Münsterberg 2001; Udolph et al. 1998). An early ventral ectoderm role for EGFR in sea urchin embryogenesis may exist during the epithelial-mesenchymal cell fate transition, specifying primary mesenchyme cells from originally ventral (vegetal) micromeres at gastrulation (Röttinger, 2004). Given these early expression patterns and/or conserved roles for EGFR homologs in *Drosophila melanogaster*, vertebrates and sea urchin, it is somewhat surprising not to see *TER-a* transcription in the *Tetranychus urticae* germdisc, when dorso-ventral axis elements would be emerging. However, dorso-ventral pattern in *Tetranychus* may arise from an initially radially symmetrical primordium through other mechanisms (e.g. putative TGF- β signalling), with potential *TER*-mediated signalling (ventral or otherwise) deployed later in development (Akiyama-Oda and Oda 2003).

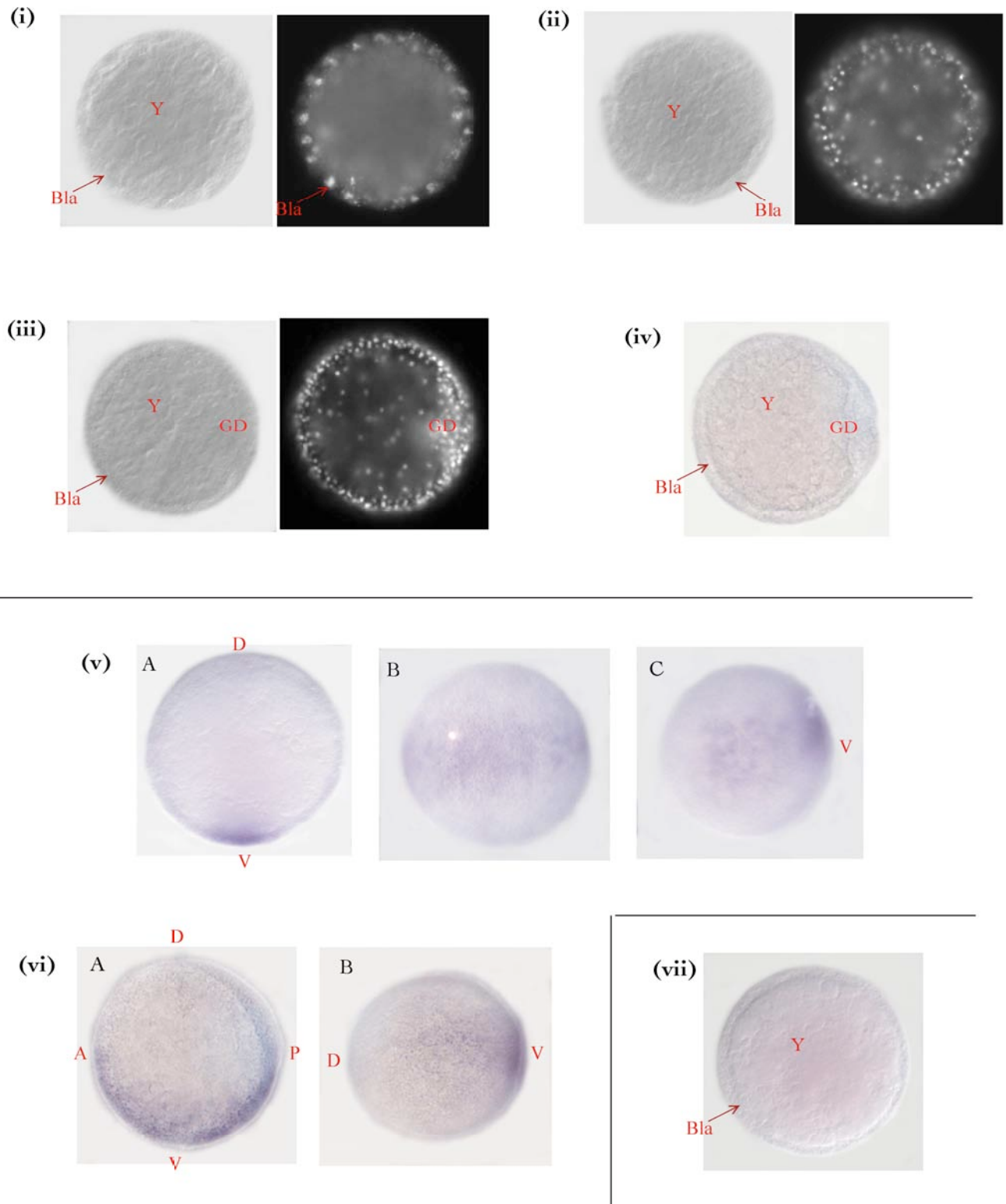


Figure 5.5.1 Detection of *TER-a* mRNA transcripts by *in situ* mRNA hybridisation on whole mount embryos. No transcription at (i) blastoderm stage, (ii) late blastoderm or (iii)-(iv) germ disc stages. Broad ventral *TER-a* domain appears at (v)-(vi) early germ band accumulation stage (14-16hr AEL). (vii) Negative control.

D - dorsal; V - ventral; A - anterior; P - posterior; Bla - blastomere cell layer; GD - germ disc; Y - yolk

5.5.2 *TER-a* at germband stage: onset of transcription

i) Early germband stage

TER-a mRNA transcripts are first detected between ~16-19hrs AEL, as a broad ventral stripe 22-25µm wide at its widest point, tapering at the anterior and posterior with presumptive dorsal and dorso-lateral regions devoid of gene activity (Figure 5.5.1,v-vi). The stripe domain corresponds to ventral ectoderm of the extending germband, and hence it may be that, as in *Drosophila melanogaster*, EGFR signalling is conferring ventral ectodermal fate as the antero-posterior axis of the spider mite germband is established. In *Drosophila melanogaster* this early role is allied to neuroectoderm patterning and, slightly later, to limb specification. Parallels have been drawn between EGFR signalling from the *Drosophila melanogaster* ventral midline and from the vertebrate floor plate, invoking multiple interactions with proneural genes and negative regulation of BMP signalling factors (Fleming et al. 1990; Kretzschmar et al. 1999; Udolph et al. 1998). The appearance of *TER-a* transcripts in a ventro-medial domain in the early *Tetranychus urticae* germband may be indicative of conserved roles for EGFR signalling in arthropods or more broadly, in both protostomes and deuterostomes: Firstly, a role in assigning ventral identity to germband cells perpendicular to the antero-posterior axis may be conserved at least in arthropods, and secondly a role in defining and mediating neuroectoderm patterning may be conserved since protostome and deuterostome divergence (De Robertis, 1996; Röttinger, 2004; Tickle, 2003; Bridge, 2000).

ii) Mid-late germband stage

Detection of TER-a transcripts

During mid to late germband development (20-22hrs AEL) the broad earlier stripe retracts from more lateral cells to form a thinner band ~10µm wide running along the entire length of the ventral midline (Figure 5.5.2,i-ii). The expression domain resembles a ring encompassing approximately 90% of the egg circumference when viewed in sagittal section, as the germband embryo extends by curving around the yolk such that posterior and anterior ectoderm regions are directly opposed on the dorsal apex (Figure 5.5.2,iiA).

Detection of TER-a + Tu-Dll transcripts

I carried out double *in situ* hybridisation with *TER-a* + *Tu-Dll* probes in late germband-early limb bud transition stage embryos (20-22hrs AEL), the *Tu-Dll* ssRNA probe synthesised from a ~800bp *Tu-Dll* insert in pGEM®-T Easy vector (c.f. Figure 3.1.4a). A combined DIG-AP phosphatase-chromophore reaction showed the ventral band of *TER-a* expression running between lateral pairs of

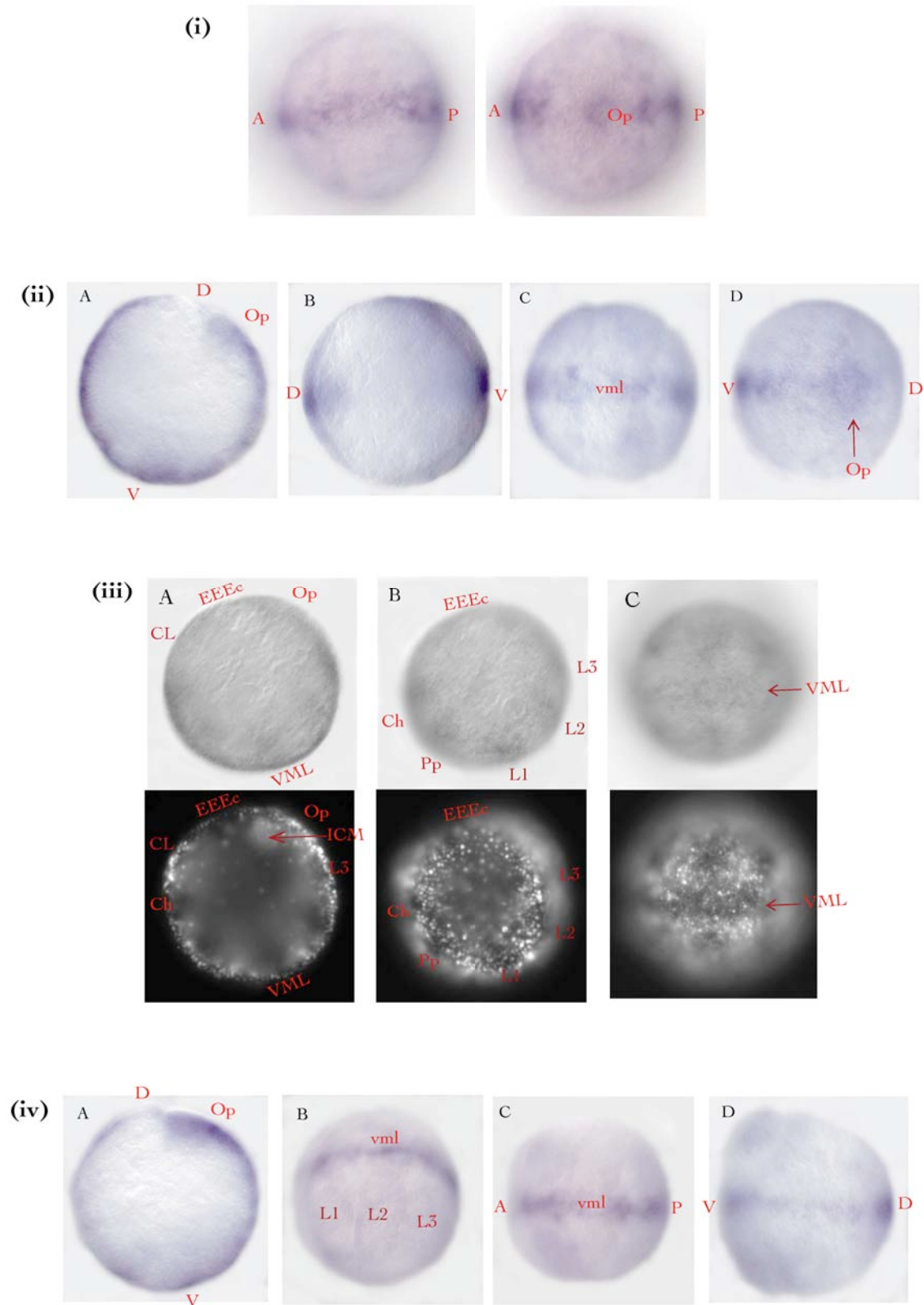


Figure 5.5.2 *TER-a* *in situ* mRNA hybridisation in (i)-(ii) late germband stage embryos. Gene expression detectable throughout ventral midline, domain expanded around the opisthosomal/posterior growth zone. (iii) Double *in situ*: *TER-a* and *Tu-Dll* mRNA transcripts detected with the same chromophore in late germband stage embryo. *TER-a* activity at ventral mid line, *Dll* activity associated with limb primordia. (iv) *TER-a* mRNA transcripts at early limb bud stage (22-25hr AEL), in narrowed ventral midline domain.

A - anterior, P - posterior, D - dorsal, V - ventral, CL - cephalic lobe; Op - opisthosoma, VML - ventral midline; EEEc - extra-embryonic ectoderm; L - walking leg; PP - pedipalp; Ch - chelicera; ICM - inner cell mass/endoderm.

what I infer to be *Dll*-expressing limb primordia (Figure 5.5.2,iii). *Dll* mRNA-expressing cells are separated from the ventral midline stripe of presumed *TER-a* positive cells by $\sim 7.8\mu\text{m}$ (~ 6 cell diameters), suggesting either absent or indirect, medium-long range regulatory interactions between the two genes. EGFR activation in *Drosophila melanogaster*, *C. elegans* and vertebrates tends to be restricted to within 1-2 cell diameters of the signalling ligand source, *via* various regulatory mechanisms (reviewed in (Shilo 2005)). Although TER-a protein domains may hypothetically extend laterally from the ventral *TER-a* transcription 'ring' domain, this is speculative given a lack of antibody data for activated members of the signalling pathway (see section 5.6) and furthermore, it is unlikely given the conserved short-range nature of EGFR signalling in other species (Shilo 2003; Shilo 2005; Sternberg and Horvitz 1991).

Comparative EGFR regulation in ventral/ventro-lateral ectoderm

In *Drosophila melanogaster*, DER is required between 4.5 and 5.5hrs AEL to further pattern ventro-medial neurogenic cells and to specify ventro-lateral limb primordia within the ventral ectoderm, parallel processes in vertebrate floor plate and limb bud development notably also involving EGFR (Kubota et al. 2000; Omi et al. 2005; Raz and Shilo 1993; Tickle 2003; Tickle and Münsterberg 2001). EGFR-induced kinase activity in *Drosophila melanogaster* mediates ventral *pp* and *Dll* repression, restricting expression to dorsal and medio-lateral focal domains respectively (Angelini and Kaufman 2005b; Kubota et al. 2000; Panganiban 2000; Raz and Shilo 1993). A similar repressive regulatory interaction between TER-a signalling and *Tu-Dll* appears not to exist in *Tetranychus urticae*; EGFR-independent regulation of the two genes during early limb specification is a more parsimonious interpretation of the double *in situ* hybridisation pattern observed.

5.5.3 Early limb bud stage: ventral neurogenic *TER-a* transcripts

TER-a gene activity at the ventral neurogenic midline

During the earliest stage of limb outgrowth from lateral ectoderm primordia (22-25hr AEL), the previous ventral *TER-a* transcription domain refines to a more narrow ($\sim 6.5\mu\text{m}$) ventral midline stripe, with a slightly more intense hybridisation signal that may indicate up-regulated *TER-a* transcription relative to the earlier germband (Figure 5.5.2,iv). The ventral midline stripe expands at the site of the posterior opisthosoma, forming an approximately diamond-shaped caudal domain (Figure 5.5.2,iiD,ivA). Double *in situ* hybridisation to complement *TER-a* with *Dll* mRNA expression domains again reveals apparently independent transcriptional domains for the two genes. *TER-a* gene activity is therefore consistently associated with the ventral midline, site of neural cell specification

and axon guidance as neural commissures are patterned and integrated (Mittmann 2002; Scholz et al. 1997).

Comparative EGFR regulation in neural patterning: hypotheses for spider mite?

Modulation of *TER-a* transcription to generate dynamic expression within the midline neurogenic domain of *Tetranychus* may depend on negative feedback mechanisms similar to those at play in *Drosophila melanogaster*. However, The nature of *TER-a* transcriptional regulation at the ventral midline and posterior germband remain hypothetical at this point; exploration and characterisation of candidate gene orthologs is needed. Considering mechanisms for negative regulation of *Drosophila* and vertebrate EGFR (summarised below) it could be appropriate to investigate putative orthologs of Argos, Kekkón, Sprouty, integrin- $\alpha_1\beta_1$ /TCPTP, IMP and Notch signal elements in *Tetranychus*:

Negative EGFR regulation in Drosophila and vertebrates

DER signalling range is delimited by negative regulators whose transcription is induced by the EGFR signalling pathway itself (i.e. an auto-regulatory loop mechanism). For example, Argos binds with the principle signalling ligand Spitz to form a heterodimer no longer able to conform with the receptor, hence repressing perpetuation of the signal beyond source midline cells (Jin et al. 2000; Klein et al. 2004; Schweitzer et al. 1995; Shilo 2005; Stemerdink and Jacobs 1997). Alternate means of negative regulation come from induction of factors that interfere with the membrane-bound receptor (e.g. Kekkón), kinase cascade elements (e.g. Sprouty) or affect ligand sequestration in the ER (reviewed in (Jorissen et al. 2003) and (Shilo 2005))(Reich, 1999; Klein, 2004; Klämbt, 2001). Recent findings show that EGFR may be also down-regulated independently of EGFR signal activation *per se*, such as by a specific T-cell protein tyrosine phosphatase (TCPTP) activated by integrin- $\alpha_1\beta_1$ in the cytoplasm, and by elements of the Notch signalling pathway that are implicated in reciprocal feedback control (de Celis et al. 1993; Fleming et al. 1990; Lai 2004; Matilla et al. 2005; Sugiura and al 2003; Wennerberg and al 2003). Argos is not known to be conserved in vertebrate EGFR signalling networks, but homologs of the inhibitors of Kekkón and Sprouty are known, suggesting that similar *EGFR* transcriptional control mechanisms may be conserved outside *Drosophila* and conceivably - due to its relative phylogenetic proximity as an arthropod – in *Tetranychus urticae* (Reich et al. 1999; Shilo 2005).

Lack of TER-a activity in early limb buds: comparative considerations

It appears that during this earliest stage of limb outgrowth in *Tetranychus*, EGFR signalling is inactive. In *Drosophila melanogaster* EGFR activity is required, after its role in specifying the 20-30 cells of the limb primordial cluster, to specify leg vs. wing primordial identity within the developing limb imaginal disc (Kubota et al. 2000; Panganiban 2000). As no notable role in initiating or directing early outgrowth has been documented for DER, the absence of limb-associated *TER-a* expression at

this stage may reflect a homologous phase of EGFR down-regulation or alternative mechanism for inducing outgrowth.

By contrast, in chick, initial limb patterning is induced by EGFR signalling *via* EGF and TGF- α , EGF activity itself dependent on IGF-I receptor signalling at axial and lateral plate mesoderm. Outgrowth depends on reciprocal and synergistic interactions between distal apical ectodermal ridge (AER) and underlying mesoderm cells (sub-ridge mesenchyme), proliferation mediated largely by FGF, EGF/TGF- α - and Wnt signalling (Dealy et al. 1998; Lax et al. 1988; Tickle 2003). As neither *Drosophila* nor *Tetranynchus* display deployment of EGFR activity in a similar context (i.e. early limb buds), EGF signalling in early limb buds is evidently a vertebrate regulatory innovation.

5.5.4 *TER-a* transcription during late limb bud development

Continued *TER-a* activity in ventral neurogenesis

Expression of *TER-a* mRNA persists at the ventral midline during late limb bud development (25-30hrs AEL) as a band narrowing from 5.6 μ m to ~2.6 μ m (Figure 5.5.3,i-ii). This suggests further refinement of signalling range by down-regulation from lateral cells adjacent to the midline, perhaps by processes orthologous to EGF auto-regulatory or independent mechanisms mentioned in 5.5.3 above.

Within the developing neurogenic mid line, *TER-a* presumably promotes further differentiation and selected neural cell survival - in keeping with the general function of EGFR signalling in vertebrates and arthropods to promote differentiation and inhibit apoptosis (Dumstrei et al. 1998; Dumstrei et al. 2002; Omi et al. 2005; Shilo 2003). In *Drosophila*, the principle ligand Spitz activates EGFR at the midline, MAPK signalling effecting midline glial and later commissural sheath development (Mayer and Nusslein-Volhard 1988; Rutledge et al. 1992; Scholz et al. 1997; SDB 2005; Stemerink and Jacobs 1997). Combining this hexapod data with evidence from the phylogenetically disparate spider mite, a function for EGFR in mediating ventral neurogenesis appears to be conserved in the arthropods.

Onset of *TER-a* activity in developing limbs

TER-a transcripts are first detected in appendage domains during the later phase of limb bud elongation, between approx. 27-30hrs AEL. mRNA expression in the limbs appears in a medial patch or pervasive medial-proximal domain, excluded from or expressed at a much lower level in distal-most domains (Figure 5.5.3,ii). Composite *TER-a* and *Tu-Dll* *in situ* hybridisation reveals transcripts throughout the distal, medial and proximo-medial limb, assumed to reflect a combination of distal *Tu-Dll* and more medial-proximal *TER-a* expression domains. The presence or degree of any overlap in

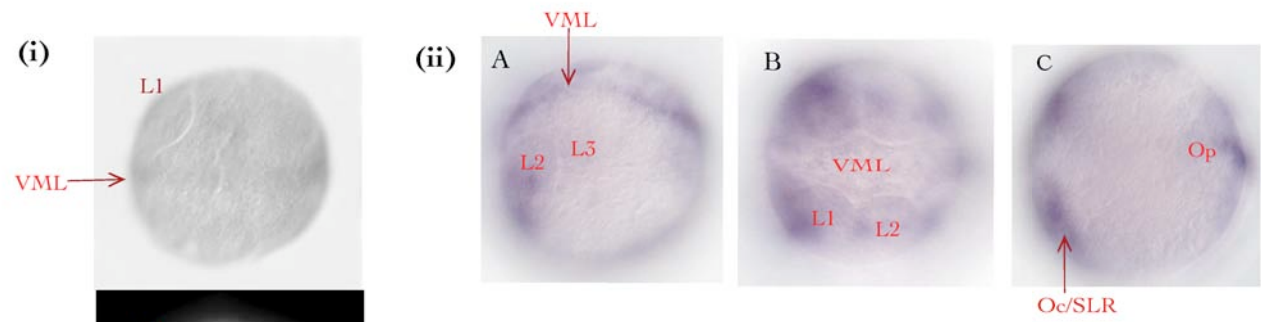
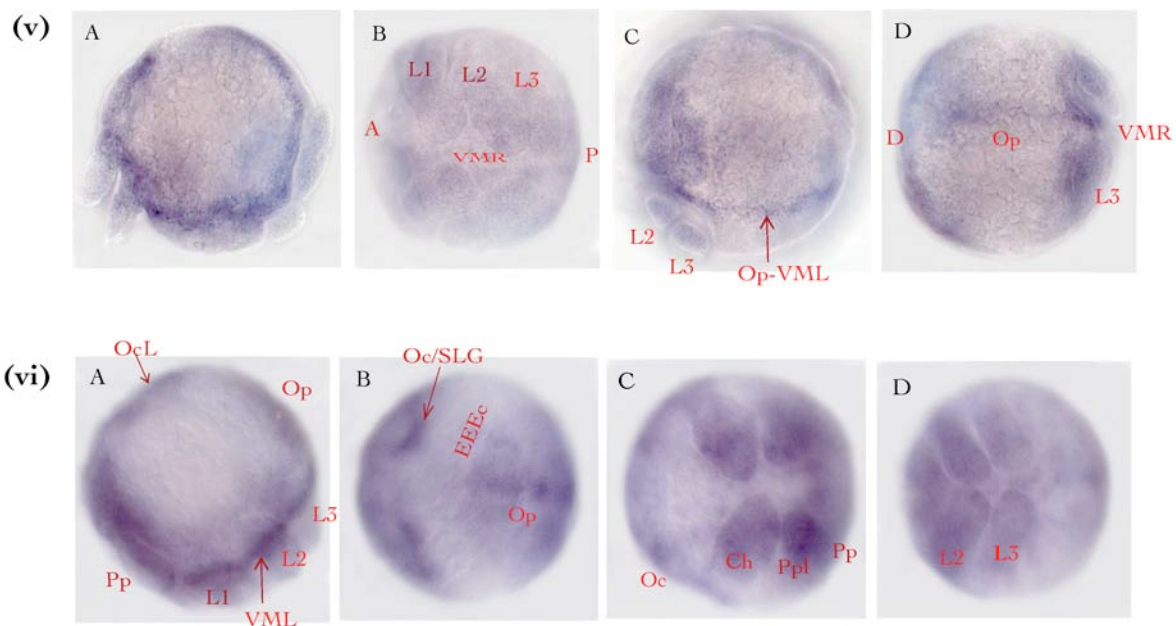
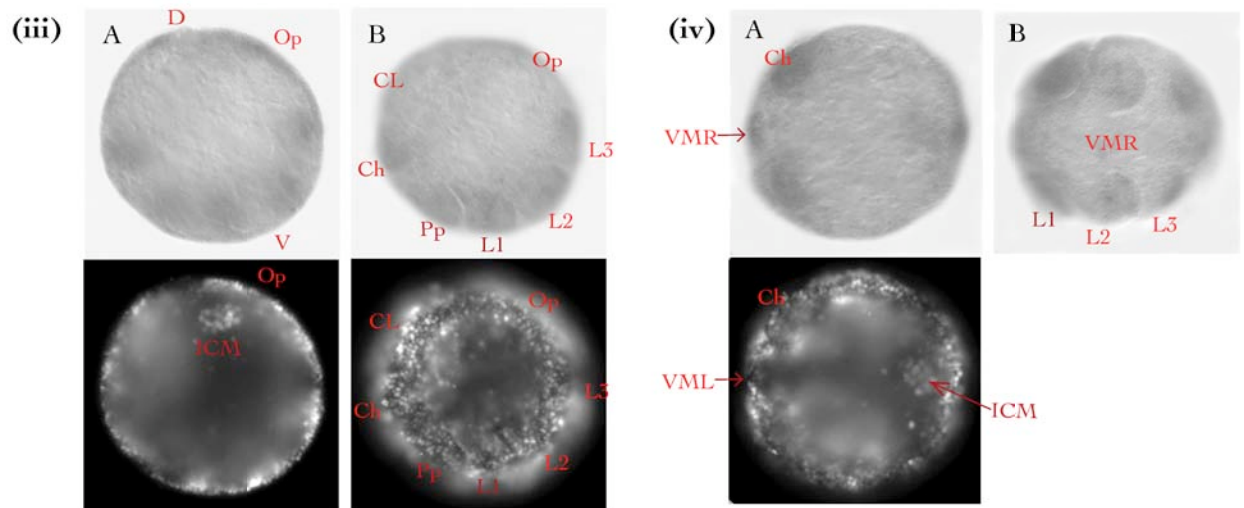


Figure 5.5.3 *TER-a* mRNA *in situ* hybridisation in (i)-(iv) late limb bud stage (25-30hr AEL) and ventral ridge contraction (v)-(vi) stage (30-39hr AEL) embryos. (i) *TER-a* gene expression in vental mid-line domain and slightly later (ii) also within subsets of appendage derived cells. (iii)-(iv) *TER-a* and *Tu-Dll* mRNA activity in ventral mid-line and appendage cells. (v)-(vi) *TER-a* transcription during ventral contraction (marked by ridge), detected at ventral midline and ocular lobe, lower levels in appendages.



CL - cephalic lobe; OcL - ocular lobe; Oc/SLG - ocular/semi-lunar groove; Ch - chelicera; Pp - pedipalp; Ppl - pp lobe; L - walking limb; Op - opisthosoma; VML/R - ventral midline/ridge; ICM - inner endodermal cell mass.

the domains cannot be determined without employing a double *in situ* hybridisation protocol with distinct chromophore-labelled mRNA probes; unfortunately, foreseeable technical and signal:noise problems, as well as time constraints prevented this from being possible.

EGFR/RTKs in proximo-distal limb patterning: comparative considerations

Regarding the interpretation of *TER-a* transcription in appendages, a comparison of *Drosophila* and vertebrate proximo-distal patterning mechanisms is worthwhile. RTK deployment has been suggested to be a conserved feature of proximo-distal limb development in both *Drosophila melanogaster* and vertebrates, mediated by a distal tip source of EGFR and FGFR signalling respectively (Campbell 2002; Dubrulle et al. 2001; Dudley et al. 2002; Galindo et al. 2002; Tickle 2003). In *Drosophila*, Wg, Dpp and Dll activate secretion of the secondary EGFR ligand Vein (Vn) at the centre of the 1st instar imaginal leg disc, corresponding to distal limb territory. By the 3rd instar, *vn* transcription is independent of Wg and Dpp signals, and a graded response to EGFR regulates differential gene expression along the proximo-distal axis, in combination with *Dll* and medial *pac* activity (Campbell 2002; Campbell et al. 1993; Galindo et al. 2002; SDB 2005). In comparison, proximo-distal patterning of vertebrate appendages is achieved by various FGFs emanating from the AER at the limb apex, signalling through another tyrosine kinase receptor, FGF-R. Positional identity may be conferred by multiple factors, including cellular location behind an AER-associated FGF gradient and direct or indirect FGF regulation by an oscillating developmental clock (Dealy et al. 1998; Delfini et al. 2005; Dubrulle et al. 2001; Tickle 2003; Tickle and Münsterberg 2001).

TER-a and conserved proximo-distal limb patterning?

Lack of apparent EGFR activation in *Tetranychus* distal limb tips casts doubt on the validity of interpreting deep homology in patterning mechanisms based on only fly and vertebrate data (Jenner 2004). Possible explanations, that cannot be assessed with these data alone, include:

- i) that the distal signalling role is absent in *Tetranychus urticae* limb development and evolved independently in lineages leading to *Drosophila*,
- ii) that alternative genetic interactions establish proximo-distal positional identities,
- iii) or that the presently observed pattern presents an exceptional modification of an original and indeed deeply conserved proximo-distal patterning mode.

Present understanding of arthropod phylogeny indicates that chelicerates are basal relative to insects, and chelicerates form limbs by direct outgrowth during embryogenesis rather than *via* imaginal discs as in *Drosophila*, although progressive outgrowth is the norm for most insects (Anderson 1973; Angelini and Kaufman 2005b; Prpic et al. 2003). As both *Tetranychus urticae* and *Drosophila melanogaster* are derived species within their respective sub-phyta, wider taxonomic sampling is desirable prior to making conclusions about the ancestral state of EGFR regulation during arthropod

proximo-distal limb development. If a lack of distal tip signalling (as in *Tetranychus*) proves to be the rule rather than the exception, it would call into question the hypothesis of conserved RTK signalling in the limb P/D axis since the last common ancestor of arthropods and vertebrates - i.e. since the divergence of protostomes and deuterostomes.

5.5.5 *TER-a* transcription during ventral ridge contraction/limb elongation

TER-a activity in later neurogenesis

Between 30-39hrs AEL significant limb elongation occurs, concomitant with ventral ridge reduction, and *TER-a* transcripts continue to be detected at a high level in a thin ventral midline stripe (2.6-2.8µm wide) that expands to a broader domain overlying the posterior opisthosoma and in the anterior direction, into the cephalic neuroectoderm (Figure 5.5.3,v-vi). Sustained *TER-a* gene activity in developing neural tissue may be compared with later neural roles for EGFR signalling in *Drosophila*, including medial brain, visual system and stomatogastric nervous system (SNS) differentiation from previously EGFR-positive medial head ectoderm cells (Dumstrei et al. 1998; Dumstrei et al. 2002; Ip et al. 1992; Kumar and Moses 2001; Kumar et al. 1998). As mentioned previously, relatedness between fly and vertebrate EGFR signalling has been proposed, as both forms promote differentiation and prevent apoptosis, as well as being implicated in negative Notch pathway regulation (Dumstrei et al. 1998; Dumstrei et al. 2002; Fleming et al. 1990; Shilo 2003). This broad hypothetical conservation is supported by the later cephalic and ventral midline neuroectoderm EGFR expression observed in *Tetranychus*.

TER-a activity during appendage morphogenesis

EGFR activity in elongating limbs appears to be down-regulated to a lower level at this stage, and detectable *TER-a* mRNA expression domains become more homogeneous, occupying all but the most proximal limb segments (Figure 5.5.3,v). This pattern suggests that EGFR has a general mitogenic role in the rapidly growing limbs, perhaps sustaining a high rate of cellular proliferation. *TER-a* expression here might reflect a comparable phase in fly limb development allied to muscle/tendon differentiation, although the homogeneity of the expression pattern weakens this interpretation, and a general role in promoting growth after distal gradient mediated proximo-distal identity specification, is not documented in *Drosophila* (Gabay et al. 1997; Schweitzer and Shilo 1997; Shilo 2003; Shilo and Raz 1991). Thus, EGFR expression between these representative, albeit specialised, chelicerate and insect species appears not to be conserved during later limb morphogenesis and differentiation. By contrast, EGF signals are deployed widely in late vertebrate limb buds, with roles in repressing chondrogenesis, myogenesis, and antagonising BMP signalling (Dealy et al. 1998; Kretzschmar et al.

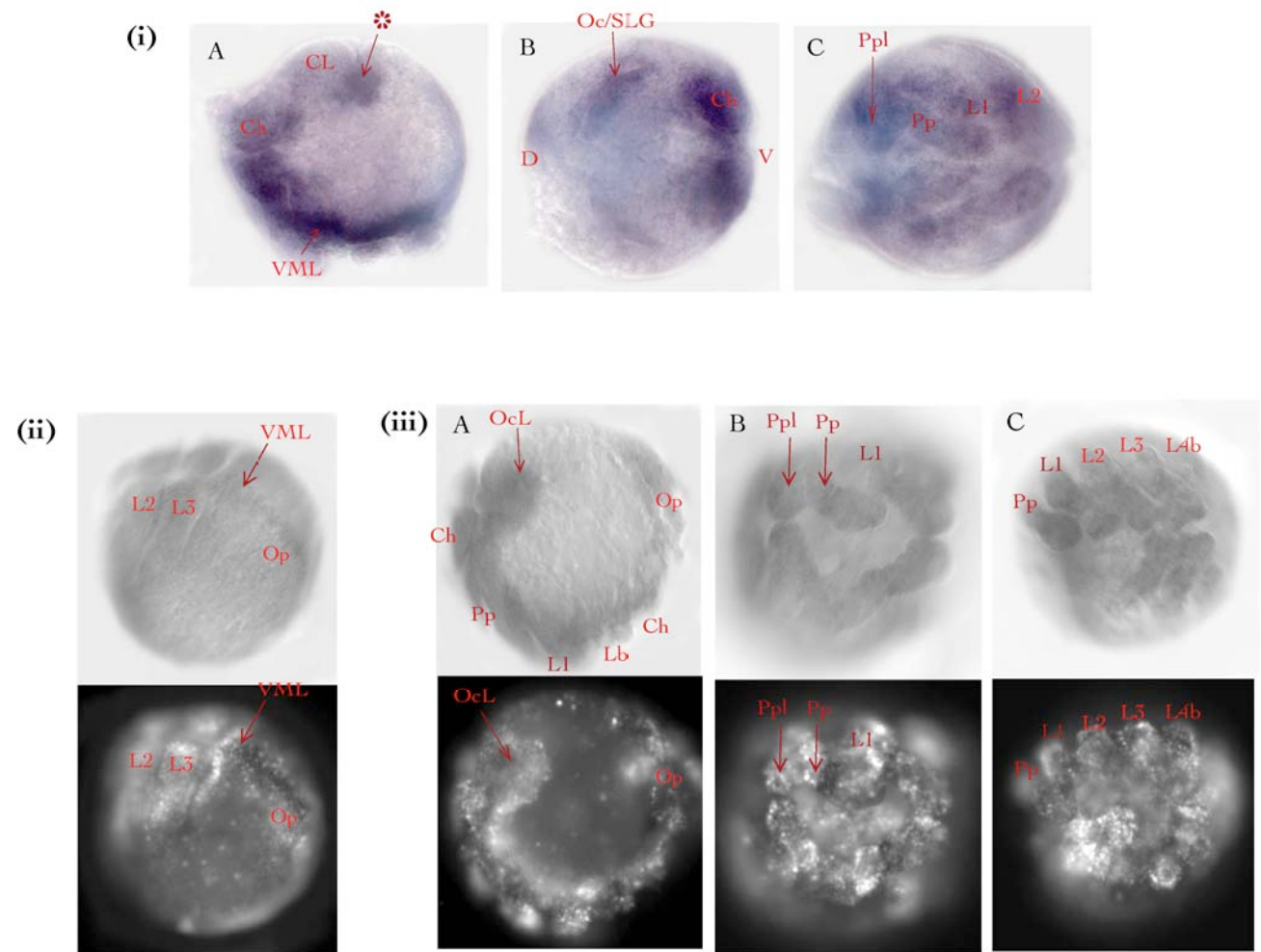


Figure 5.5.4 *TER-a* single and *TER-a/Tu-Dll* double *in situ* mRNA patterns in ventral closure, germband contraction and prelarva stage embryos.

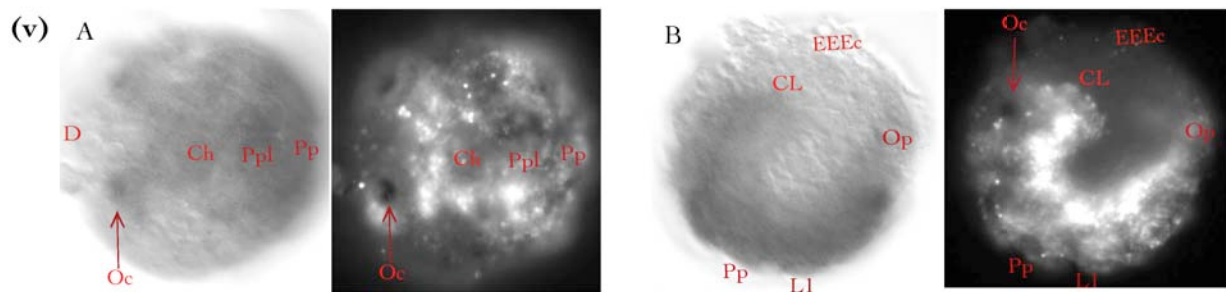
(i)-(ii) *TER-a* expression at ocular/cerebral grooves and putative cerebral tissue (asterisk), ventral midline cells and only at low level in appendages, esp. proximally and in proximo-distal patches on L1-3.

(iii) *TER-a/Tu-Dll* joint transcription pattern during early germband contraction: more appendage-specific gene activity further to *TER-a*-only pattern, especially *Dll*-diagnostic medial rings.

(iv) Single *TER-a* in situ at early germband contraction: strong transcription reported at VML and ocular/cerebral domains.

(v) Dorsal closure/pre-larval stage *TER-a/Tu-Dll* double in situ: gene activity clearest in ocular/cephalic lobe and throughout appendages.

Oc - ocular segment; Oc/SLG - ocular/semi-lunar groove; CL - cephalic lobe; Ch - chelicera; Pp - pedipalp; Ppl - Pp lobe; L - walking limb; Op - opisthosoma; EEec - extra-embryonic ectoderm; VML - ventral midline; D - dorsal.



1999; Tickle 2003). As such, a general role for EGFR signalling in proximo-distal appendage patterning might possibly have existed in the body wall outgrowths of an arthropod-vertebrate common ancestor, with variation in transcriptional dynamics occurring *via* disparate regulatory evolution after lineage separation.

5.5.6 *TER-a* transcription during ventral closure

TER-a transcriptional activity is detected in three general domains in early and later ventral closure stage embryos (40-45hr AEL), largely persisting from previous developmental phases: a ventral midline strip, differentiating cephalic neural tissue, and localised within the rapidly elongating appendages.

TER-a in ventral neural differentiation

TER-a positive cells continue to form a narrow band, estimated at 3.5-3.8µm wide, running the length of the ventral midline, right to the end of the posterior opisthosoma (Figure 5.5.4,i-iii). As mentioned previously, EGFR signalling from the VML is likely to be associated with ventral neurogenesis, as well as modulating cell survival and apoptosis within differentiated structures (SDB 2005; Stemerdink and Jacobs 1997; Urban et al. 2004; Yagi et al. 1998). The slight increase in width of the midline stripe may be an artefact from pseudo-spherical object measurement, or it may indicate a small lateral retreat of negative regulatory factors, expanding the domain of mitogenic EGFR activity.

TER-a modulation in anterior neural differentiation

Strong *TER-a* expression is observed in semi-lunar grooves, associated with developing optic lobes and cerebrum (Figure 5.5.4,iB,iiiA). Differentiation and invagination of previously specified cephalic neuroectoderm cells is evident from dynamic head lobe ectoderm expression and appearance of a defined population of *TER-a* positive cells in a hypothetically medial brain location (Figure 5.5.4,iA,iB,iiiA). These domains are consistent with later roles for EGFR in directing cephalic neuroectoderm development, mediating differentiation of the visual system and brain, and affecting cellular adhesion properties to permit invagination. In *Drosophila melanogaster* EGFR signalling is also required to direct differentiation of the medial brain, visual system and stomatogastric nervous system (SNS) from precursor embryonic head midline ectodermal cells (Dumstreit et al. 1998; SDB 2005). It has been found that EGFR partly modulates optic placode morphogenesis *via* DE-cadherin (encoded by *shotgun*, *shg*), which promotes cell survival and adhesion (Dumstreit et al. 2002). Commonalities between *Tetranychus urticae* and *Drosophila melanogaster* EGFR deployment indicate support for a conserved role in arthropod anterior neuroectoderm development.

TER-a in late appendage morphogenesis

TER-a is expressed throughout the extending limbs, with specific regions of relatively up-regulated activity located in transverse distal and medial rings or patches. As previously suggested, EGFR activity in the appendages may correlate with promoting general proliferation, limb morphogenesis and the differentiation of muscle/tendon attachment sites close to joints. Indeed, near-transverse or restricted lateral *TER-a* expression domains along the proximo-distal axis support a link with muscle and articulated joint development. Considering the demonstrated role for EGFR signalling in defining muscle and tendon precursors and attachments in *Drosophila melanogaster*, the data for *Tetranychus* infer a common function for EGFR in late arthropod limb morphogenesis (Dealy et al. 1998).

5.5.7 *TER-a* transcription during germband contraction and dorsal closure

Germband contraction, dorsal closure (45-50hrs AEL) and late limb patterning events are accompanied by continued *TER-a* expression at the neurogenic ventral mid-line (VML), and strong *TER-a* activity in developing cerebral tissues – manifest as a broad domain encompassing invaginated cerebral ganglia and optical system elements (Figure 5.5.4,iv-v). In addition, only low level *TER-a* expression is observed along proximo-distal limb axes, except in limited cases where I detected transcript signal at distal limb tips.

TER-a in ventral midline and cerebral neurogenesis

Persistence of ventral midline expression (Figure 5.5.4,iv) presumably indicates a continued function for EGFR in selecting cells for survival during differentiation of the ventral nerve ganglion network, representing a conserved role shared with *Drosophila* and possibly more disparate, non-arthropod taxa (Dealy et al. 1998; Scholz et al. 1997; Stemerink and Jacobs 1997). The cerebral *TER-a* domain is greatly expanded relative to ventral closure stage expression, reflecting proliferation and a hypothetical role for *TER-a* in directing brain differentiation beyond the earlier, more restricted medial domain (c.f. Figure 5.5.4,i). Strong signal is detected in semi-lunar grooves of the optic lobes, presumably related again to visual system differentiation and invagination mediated by regulated intercellular adhesion affinities. EGFR signalling may play a wider role in cerebral patterning in *Tetranychus urticae* than it does in *Drosophila melanogaster*, but aside from this it would appear that the previously implied conservation of EGF signalling in arthropod anterior neurogenesis still stands.

Reduced TER-a transcription in final stages of appendage patterning

TER-a gene activity appears either considerably or completely down-regulated in appendages after 45hrs AEL (Figure 5.5.4,iv). The exception to this occurs in embryos at early stages of germband

contraction, in which putative transcription is detectable in distal tips of the chelicerae (Ch) and pedipalpal lobes (Ppl). This distal *TER-a* expression may be related to a transient phase of EGFR signalling, promoting the extensive morphogenetic changes and PNS structural differentiation that occur in the Ch and Ppl during gnathal development (Crooker et al. 2003; Jeppson et al. 1975).

Double *in situ* hybridisation with *TER-a* + *Tu-Dll* ssRNA probes carrying the same chromophore (for reasons previously explained) revealed notably darker transcription signal in prosomal limbs (Figure 5.5.4,v). As I have shown *Tu-Dll* to be expressed in developing limbs (c.f. Chapter III), I interpret the presence of strong limb staining in double *TER-a* + *Tu-Dll* but not in *TER-a* single *in situ* as indicating active *Dll* transcription but absence of (repressed) *TER-a* transcription. Further more, as *Tu-Dll* is transcribed during anterior neurogenesis, combined *TER-a* + *Tu-Dll* *in situ* hybridisations cannot provide information regarding differential anterior transcription, as the two genes are likely to be partly or wholly co-expressed (c.f. Chapter III, section 3.2).

5.6 Testing for *EGF* signalling in *Tetranychus* embryogenesis with a dpMAPK antibody

5.6.1 Rationale for using a dp-MAPK antibody

Patterns of dual MAPK phosphorylation (dpMAPK) mirror domains of EGFR activity, and a dpMAPK antibody has been successfully used to track EGFR signalling dynamics in *Drosophila melanogaster* and the sea urchin *Paracentrotus lividus* (Gabay, 1997; Kubota, 2000; Röttinger, 2004). At the cell surface, EGF ligand binding to the extracellular EGF receptor domain induces intracellular cytoplasmic activation of the G protein Ras, Ser/Thr kinase Raf, then dual phosphorylation of MEK protein (Thr/Tyr kinase) at the top of a cascade of mitogen activated protein (MAP) Ser/Thr kinases (Gilbert 2000; Hsuan et al. 1989; Jorissen et al. 2003; Kretzschmar et al. 1999; Shilo and Raz 1991). The kinase cascade ends in phosphorylation of target transcription factors (Pointed/Ets family proteins), generating a transcriptional activation output (Jorissen, 2003; Röttinger, 2004). The monoclonal antibody available specifically recognises the 11 amino acid activation loop of dpMAPK, and wherever it is detected we can infer EGF signal transduction in progress (Gabay et al. 1997).

The notable phylogenetic disparity of *Drosophila melanogaster* (Insecta) and *Paracentrotus lividus* (Echinodermata) suggested to me that the α -dpMAPK antibody targets a highly conserved epitope and hence might positively cross-react in *Tetranychus urticae* embryos, allowing visualisation of EGFR signalling activity to complement *TER-a* mRNA transcript data.

5.6.2 Lack of *Tetranychus* cross-reactivity with anti-dp-MAPK

I assayed whole mounted *Tetranychus urticae* embryos for EGF signal transduction by immunostaining with DIG-AP labelled α -dpMAPK, but detected no signal at any stage of development. I tested multiple primary antibody concentrations (from 1:10 to 1:2000), peroxidase reaction times (from 5' – 45') and reaction temperatures (from 4°C – 25°C), but embryos from blastoderm to dorsal closure stage never exhibited more than general background staining. Positive control reactions, detecting α -phospho-Histone-III revealed the normal pattern of Histone-III phosphorylation in dividing cells.

Consistently negative results indicate that either:

- i) the epitope for the dpMAPK antibody does not cross-react in *Tetranychus urticae*, or
- ii) that the specific target epitope is biochemically disrupted because of sensitivity to the embryo fixation protocol used prior to immunostaining.

As a chelicerate *Tetranychus* is widely separated in terms of evolutionary divergence from both *Drosophila melanogaster* and *Paracentrotus lividus*, animals in which positive cross-reaction with α -dp-MAPK is documented. Therefore, regarding explanation (i), MAPK activation loop modifications could have arisen independently in residues of the spider mite lineage MAPK ortholog after its divergence from stem arthropods, followed by fixation of the 'altered' epitope so that effective binding *via* the ancestral dpMAPK epitope is prevented. Although positive control experiments were successful, explanation (ii) is also feasible as biochemical stability and sensitivity varies between any given epitopes, and α -p-His-III displays notably greater robustness to physico-chemical stress relative to many other antibodies (Extavour 2003; Khila 2006a).

Conclusions to Chapter V

i) *Tetranychus* ∂pp

In section 5.1 of this chapter we have seen that the *Tu- ∂pp* gene cloned by degenerate PCR, iPCR and cDNA library screening corresponds to a genuine spider mite $\partial pp/BMP2/4$ ortholog. The 1071bp available from gene sequencing encompasses the 5' untranslated region (UTR), complete coding sequence (CDS), 3' UTR and, in cDNA clones, a terminal ribosome drop-off site and polyadenylated tail. The spider mite Dpp pro-domain contains basic residues likely to be cleaved by subtilisin family proteases to release a mature Dpp ligand domain, and both pro- and ligand domains contain apparently conserved N-linked glycosylation sites that may be used in labelling for secretion. The TGF- β family 7x cysteine motif is perfectly conserved within *Tu-Dpp*, allowing a stable cysteine knot structure and functionally active Dpp dimers to form.

In situ hybridisation of whole mount embryos with *Tu- ∂pp* ssRNA probes did not work (section 5.2). I have suggested that this is most likely to have been due to embryonic mRNA or probe ssRNA degradation, especially as ∂pp has reportedly been found in a *Tetranychus urticae* EST project in Canada, indicating that the *Tu- ∂pp* gene must be transcribed at some point in embryogenesis and hence should be detectable. In section 5.3, I have shown that detection of Dpp signal transduction (and by inference *Tu- ∂pp* activation) with an antibody against phosphorylated Smad1 (α -p-Smad-1) is a useful approach in the hunter spider *Cupiennius salei* to complement gene expression data pertaining to *Cs- ∂pp* transcription, but unfortunately α -p-Smad-1 does not seem to cross-react in *Tetranychus urticae*. Smad-1 phosphorylation patterns are dynamic in *Cupiennius salei* development, associated with the posterior germband and, reflecting *Cs- ∂pp* expression, associated with early germband segmentation, primordial limb cell clusters, distal early limb and proximo-distal later limb patterning. The apparent activation role of *Cs-Dpp* signalling in defining early limb primordia is borne out by observations in other spiders (e.g. *Pholcus*, *Achaearanea*) and indicates a *Dll*-regulatory role highly divergent relative to that in *Drosophila* limb specification, in which *Drosophila* Dpp restricts early *Dll*-positive, limb primordial cells. Distal restriction of the Dpp morphogen upon initiation of limb outgrowth is observed in many arthropods and may be significant with respect to the nature of molecular interactions with factors such as Wingless within the antero-posterior axis of developing three dimensional appendages. In *Tetranychus*, blastoderm stage embryos seemed to show a hemispherical α -p-Smad-1 distribution, potentially allied to a conserved role for Dpp in early dorso-ventral axis specification, but background staining and poor signal detection at all other stages of embryogenesis make this conclusion speculative. Clear posterior and anterior patches of apparent α -

p-Smad-1 localisation are present from early limb bud outgrowth and ventral ridge contraction stages onwards, but I have argued that these represent artefacts caused by ingress of tissues that independently form the proctodeum and stomodeum, and indirectly capture antigen molecules. Unfortunately therefore, spider mite *∂pp* mRNA transcription patterns remain unknown at this point, limiting our ability to test hypotheses. *Tu-∂pp* gene expression, and ideally loss-of-function phenotypes, are required to clarify the nature and significance of spider mite Dpp signalling in relation to *Dll* regulation and appendage specification.

ii) *Tetranychus EGFR (TER)*

Section 5.4 of this chapter has documented successful cloning of ~1.5Kb of the *Tetranychus urticae EGFR* ortholog (termed '*TER*'), its orthology demonstrated by Bayesian inference methods. The *TER* fragment is an incomplete open reading frame clearly coding for most of the intracellular tyrosine kinase domain typical of receptor tyrosine kinases (RTKs) such as EGFR, yet structural comparison indicates that CDS for a predicted N' signal peptide, extracellular domain, transmembrane domain and C' part of the TK domain remain to be sequenced. Unravelling alongside *TER* sequence analysis we have seen the identification of an apparent second, *Tetranychus urticae EGFR* gene. This gene, initially termed '*TER-b*' shares higher sequence similarity to *TER* than the EGFR orthologs of insects such as *Anopheles gambiae* and *Drosophila melanogaster*, but could not be recovered again in any experiments after initial degenerate PCR screening. I have proposed that the putative *TER* paralog recovered during the first degenerate PCR screen most likely derives from another arthropod that contaminated the initial DNA preparation: as predatory mites sporadically appear in the *Tetranychus* culture, they are the most probable source of non-spider mite, yet apparently chelicerate-like, genetic material.

TER mRNA transcription throughout *Tetranychus* embryogenesis (section 5.5) has revealed multiple, dynamic roles for *TER* in ventral midline and neurectoderm specification, anterior cephalic neurogenesis (e.g. brain and ocular differentiation) and some aspects of proximal-medial limb patterning during early limb elongation. Comparing patterns of *Tetranychus EGFR* activity with those observed in *Drosophila melanogaster* as well as more disparate taxa such as vertebrates and sea urchins, I have suggested that EGFR-mediated signalling is broadly conserved in early dorso-ventral patterning and central nervous system development, and within Arthropoda EGFR seems to have conserved anti-apoptotic functions in selecting cells for survival and differentiation at ventral midline neurogenic ectoderm, and in guiding differentiation during morphogenesis of invaginating anterior CNS structures such as the visual system and mid-brain. Notably, *TER* mRNA expression data cast doubt on the previously published idea that the *Drosophila EGFR* gradient that patterns the limb

proximo-distal axis is homologous to the FGFR proximo-distal gradient that patterns the vertebrate limb bud, unless a functional proximo-distal RTK gradient did exist in a common arthropod-vertebrate ancestor but was secondarily replaced by another mechanism in spider mite. In spite of conserved EGFR activity in arthropod neurogenesis however, *TER* and *Tu-Dll* are expressed in entirely exclusive domains during limb specification and early limb outgrowth, indicative of independent regulation, *TER* having no, or very little, effect on *Dll* transcription. As for chelicerate *dp* activity with respect to dorsal limb specification cues, it seems therefore that chelicerate (spider mite) EGFR signalling does not play a conserved role in *Dll* regulation during ventral limb primordium specification when compared with the genetic co-ordinate network described for *Drosophila*. (No information can be added regarding patterns of MAPK activation induced by EGFR kinase activity, as section 5.6 shows that a dpMAPK antibody cross-reactive in insects and sea urchins is not cross-reactive in *Tetranychus*, most likely due to divergence in critical epitope residues.)

CHAPTER VI

REGULATION OF APPENDAGE SPECIFICATION BY CANDIDATE ANTERO-POSTERIOR PATTERNING GENES:

TETRANYCHUS UBX AND ABD-A



Pycnogonon des Baleines

Introduction to Chapter VI

Why study spider mite *Ultrabithorax* and *abdominal-A*?

As described in Chapter I, *Ultrabithorax* (*Ubx*) and *abdominal-A* (*abd-A*), are expressed in the *Drosophila* abdomen and their encoded proteins interact with the early *Dll* enhancer *via* a number of co-factors (e.g. Pbc class and segmentation proteins). *Ubx* and *AbdA* repress the early *Dll-504* enhancer, modifying the output of segmental antero-posterior and dorso-ventral limb-positioning cues such that limb primordia are entirely repressed throughout the abdominal domain of active, early *Bithorax* complex (*BX-C*) gene expression. *Ubx* and/or *abdominal-A* expression have been observed in the limbless chelicerate opisthosoma without *Dll*, which is only expressed in prosomal limb primordia. This compelling Hox-*Dll* exclusivity has invited speculation about conserved vs. convergent mechanisms of achieving limbless posterior tagma in the disparate arthropod groups Chelicerata and Insecta. Thus, in this chapter I present the results of a PCR screen for *Tetranychus* Hox genes, followed by further work on posterior class BX-C Hox genes, *Ultrabithorax* and *abdominal-A*, candidate genes in the study of arthropod *Dll* regulation. I tested for exclusive domains of *BX-C* and *Dll* gene activity in the secondarily reduced spider mite opisthosoma, aiming to provide further information for models of conservation vs. divergence in *Ubx-Dll* and/or *AbdA-Dll* genetic interactions during limb specification between chelicerates and other major arthropod lineages, relative to the *Drosophila* scenario.

Tetranychus Ubx

In section 6.1.1 I report cloning of a short (~180bp) fragment of the *Tetranychus Ubx* (*Tu-Ubx*) homeobox during a degenerate PCR screen, in which a broad range of other anterior, central and posterior class Hox genes were also identified. Inverse PCR allowed extension of *Tu-Ubx* sequence to ~900bp, including the full coding sequence with conserved homeobox and *Ubx*-specific signature motifs as shown. Section 6.2 presents *Tu-Ubx* mRNA transcription patterns throughout embryogenesis, revealing a stable domain of gene activity localised to the opisthosoma, with an straight anterior boundary apparently at the Op2 segment and a posterior boundary in front of the telson or posterior-most segments. The *Tu-Ubx* expression domain tapers in line with opisthosomal morphology, and in late embryogenesis appears to be down-regulated at the ventral midline and contracts laterally and axially in accord with extreme opisthosomal segment reduction. I show in section 6.3 that *Tetranychus* embryos do not cross-react with an antibody against a conserved *Ubx*-

AbdA epitope, and analyse this failure in the light of onychophoran and arthropod protein Ubx sequence comparisons. Adding *Tu-Ubx* coding sequence and mRNA expression data to current information on arthropod *Ubx*, *abdA* and *Dll* interactions I consider the possibility that distinct genetic mechanisms for BX-C-mediated *Dll* repression evolved independently in chelicerate and insect lineages, responsible for convergent appearance of limbless posterior segments.

Tetranychus abdA

Section 6.1.4 documents extensive attempts to amplify a *Tetranychus abdominal-A* ortholog by degenerate PCR screens of genomic DNA and cDNA, none of which were successful. I used degenerate primers against conserved Hox homeobox residues and against a conserved N' hexapeptide motif that has roles in Hox-Hth-Exd trimeric complex formation, trimer-DNA binding stability and specificity, and affecting target repression or activation *via* cofactors and a conserved linker region between the hexapeptide and homeodomain. I propose that the chelicerate *abdA* ortholog has been lost in the spider mite lineage in association with extreme posterior reduction of the acarid body plan. This hypothesis is supported by absence of *abdA* in another mite (*Archegozetes longisetosus*) whose opisthosoma is secondarily reduced, and independent loss of *abdA* in the rhizocephalid crustacean *Sacculina carcini*, which has a reduced, vestigial abdomen.

6.1 Cloning and analysis of *Tetranychus Hox* genes

6.1.1 Amplification of *Tetranychus Hox* orthologs by degenerate PCR

i) *Tu-Hox* PCR conditions

In screening genomic *Tetranychus urticae* DNA by degenerate PCR, I amplified putative *Tetranychus* homologs of multiple anterior, central and posterior class *Hox* genes, as well as the homeobox gene *even-skipped*. Degenerate primers were designed to target conserved Hox protein N'/C' homeobox motifs: Hox-F1 against LELEKEF (5'-CG GAT TCC CTA GAG CTN GAR AAR GAR T-3'); Hox-R1 against IWFQNKRMN (5'-GGA ATT CAT ICK ICK RTT YTG RAA CCA IAT YT-3'); and Hox-R2 against WFQNKR (5'-G CTC TAG ACG ICG RTT TTG RAA CCA-3'). Primer sequences are based on those from Cook et al. (2001) and Martinez et al. (1997), and are listed in Appendix 2.7 (Cook et al. 2001; Finnerty and Martindale 1997; Martinez et al. 1997). *Hox* gene amplification was optimised against homeobox-containing non-target gene amplification by a PCR program with early cycles of ascending annealing temperature to gradually increase binding specificity, and augmenting standard PCR reactions to 2.5mM [Mg²⁺] to aid molecular kinetics by promoting ion-mediated DNA helix separation.

ii) Identification of *Tu-Hox* genes

I sequenced multiple clones of spider mite *Hox*-related (*Hox-r*) genes, yielding partial homeodomain coding sequence for *labial*-related (x2), *proboscipedia-r*, *Deformed-r*, *Sex combs reduced-r*, *Antennapedia-r* (x4), *Ultrabithorax-r* and *Hox10-r* (= *AbdB*-like), as well as one *Hox* 'central class' and one *Hox* 'posterior class' gene that could not be more specifically identified (Figure 6.1.1a). I assigned putative homologies on the basis of BLAST sequence identity and the presence of gene-specific signature motifs or residues, as inferred from comparisons of Hox class proteins from diverse animal phyla (Altschul et al. 1990; Cook et al. 2004; Cook et al. 2001; de Rosa et al. 1999). These signature amino acid residues occur between helices I - III and are highlighted in Figure 6.1.1b, an alignment of *Tetranychus urticae* Hox and Even-skipped homeodomains relative to a *Drosophila melanogaster* reference sequence (de Rosa et al. 1999; Hughes and Kaufman 2002b).

iii) Phylogenetic analysis of *Tu-Hox* sequences

I carried out Bayesian analysis of a 180bp homeobox region, comparing *Tetranychus Hox* genes with orthologs throughout the *Hox* cluster of various protostomes and deuterostomes. The resulting

Figure 6.1.1a
Page 1 of 2

<i>Lab-rA</i>	91/31 CAG GTT AAG ATC TGG TTC CAG AAC CGC CGC ATG gln val lys ile trp phe gln asn arg arg met 121/41
<i>Lab-rB</i>	CAG GTT AAG ATC TGG TTC CAA AAC CGC AGC ATG gln val lys ile trp phe gln asn arg ser met
<i>Pb-r</i>	CAG GTT AAA GTT TGG TTT CAA AAC CGC CGT CTA gln val lys val trp phe gln asn arg arg leu
<i>Hox-CC</i>	CAG ATA AAA ATC TGG TTT CAA AAC AGC CGC ATG gln ile lys ile trp phe gln asn ser arg met
<i>Dfd-r</i>	CAG ATA AAA ATC TGG TTC CAA AAC AGC CGC ATG gln ile lys ile trp phe gln asn ser arg met
<i>Scr-r</i>	CAG ATC AAA ATC TGG TTC CAA AAC AGC CGC ATG gln ile lys ile trp phe gln asn ser arg met
<i>Antp-rA1</i>	CAA ATT AAG ATC TGG TTC CAA AAC CGC AGC ATG gln ile lys ile trp phe gln asn arg ser met
<i>Antp-rA2</i>	CAA ATT AAA ATC TGG TTT CAA AAC CGC CGC ATG gln ile lys ile trp phe gln asn arg arg met
<i>Antp-rA3</i>	CAG ATA AAA ATC TGG TTT CAA AAC AGC CGC ATG gln ile lys ile trp phe gln asn ser arg met
<i>Antp-rB</i>	CAA ATT AAA ATC TGG TTT CAG AAC CGC CGC ATG gln ile lys ile trp phe gln asn arg arg met
<i>Ubx-r</i>	CAG ATT AAG ATC TGG TTC CAA AAC CGC CGC ATG gln ile lys ile trp phe gln asn arg arg met
<i>Hox-PC</i>	CAG GTT AAA ATT TGG TTC CAA AAC CGC CGC ATG gln val lys ile trp phe gln asn ala ala met
<i>Hox10-r</i>	CAG GTG AAG ATC TGG TTC CAA AAC CGC CGC ATG gln val lys ile trp phe gln asn arg arg met
<i>eve-r</i>	ACA ATC AAA GTT TGG TTT CAA AAC CGC CGT CTA thr ile lys val trp phe gln asn arg arg leu 91/31 121/41

Figure 6.1.1a

Comparative sequence data (123bp) for homeobox-containing genes recovered by degenerate *Hox* PCR in *Tetranychus urticae*. A number of *Hox* genes representing anterior, central and posterior class genes, as well as an even-skipped ortholog, were cloned. Multiple possible orthologs are shown for some genes (e.g. *Labial*, *Antp*), and the affinity of others is not clear (e.g. *Hox-CC*; central class; *Hox-PC*: posterior class). Given the conserved nature of the homeodomain, all orthology designations are conditional upon confirmation with extended sequence data, and genes are hence named as *Hox*-‘related’ (-r).

	...Helix1	Helix2	Helix3...	
Dm-Antp	LELEKEFH F NRYLT RRRRIEIAHA L CLT ERQIKIWFQ NRRM			Reference Hox
Lab-rA	LELEKEFH Y NRYLTRA R RRRIEIA Q SL G L N ETQ V KIWFQ NRRM			} Anterior class Hox
Lab-rB	LELEK S FHYNRY K TRARRRIEIA Q SL G L N ETQ V KIWFQ NRRSM			
Pb-r	LELEKEFL N KYL C RRRIEIA S T L EL S ERQ K W WFQNRRL			
CC-Hox	LELEKEF R FNRYL T RRRRRIEIA H ALCL T ERQIKIWFQ NRRSM			} Central class Hox
Dfd-r	LELEKEFH Y NRYL T RRRRRIEIA H SL L L S ERQIKIWFQ NRRSM			
Scr-r	LELEKEFH Y NRYL T RRRRRIEIA H SLCL S ERQIKIWFQ NRRSM			
Antp-rA1	LELEKEFH Y NRYL T RRRRRIEIA N ILCL T ERQIKIWFQ NRRSM			
Antp-rA2	LELEKEFH Y NRYL T RRRRRIEIA N ILCL T ERQIKIWFQ NRRM			
Antp-rA3	LELEKEFH Y NRYL T RRRRRIEIA N ILCL T ERQIKIWFQ NRRSM			
Antp-rB	LELEKEFH Y NRY S TRRRRIEIA N ILCL T ERQIKIWFQ NRRM			
Ubx-r	LELEKEF H T S QYL T RRRRRIE L AHALCL S ERQIKIWFQ NRRM			} Posterior class Hox
Hox-PC	V ELEKEFT V NFY T K Q RRFEL S RALCL S ERQ V KIWFQ NRRM			
Hox10-r	F ILE Q EYLM S TY T T R Q R LE L ARNLS L TERQ V KIWFQ NRRM			
eve-r	LELEKEF T A E K Y V S R P RR C ELASS L N L PE S T I K V W WFQNRRL			Even-skipped

Figure 6.1.1b Multiple sequence alignment for consecutive amino acids of *Tetranychus* homeodomain-containing proteins; 13 Hox proteins and an Even-skipped ortholog. **Blue**: divergent residue relative to *Drosophila* Antp reference sequence (a central class Hox used as reference in de Rosa et al., 1999). **Red**: divergent residue serving to diagnose particular Hox proteins, whether by specific amino acid signature or by conserved variable sites. **Boxed**: Homeodomain helices, helix 2 complete, helix 1 missing 4aa N', helix 3 missing 4aa C'.

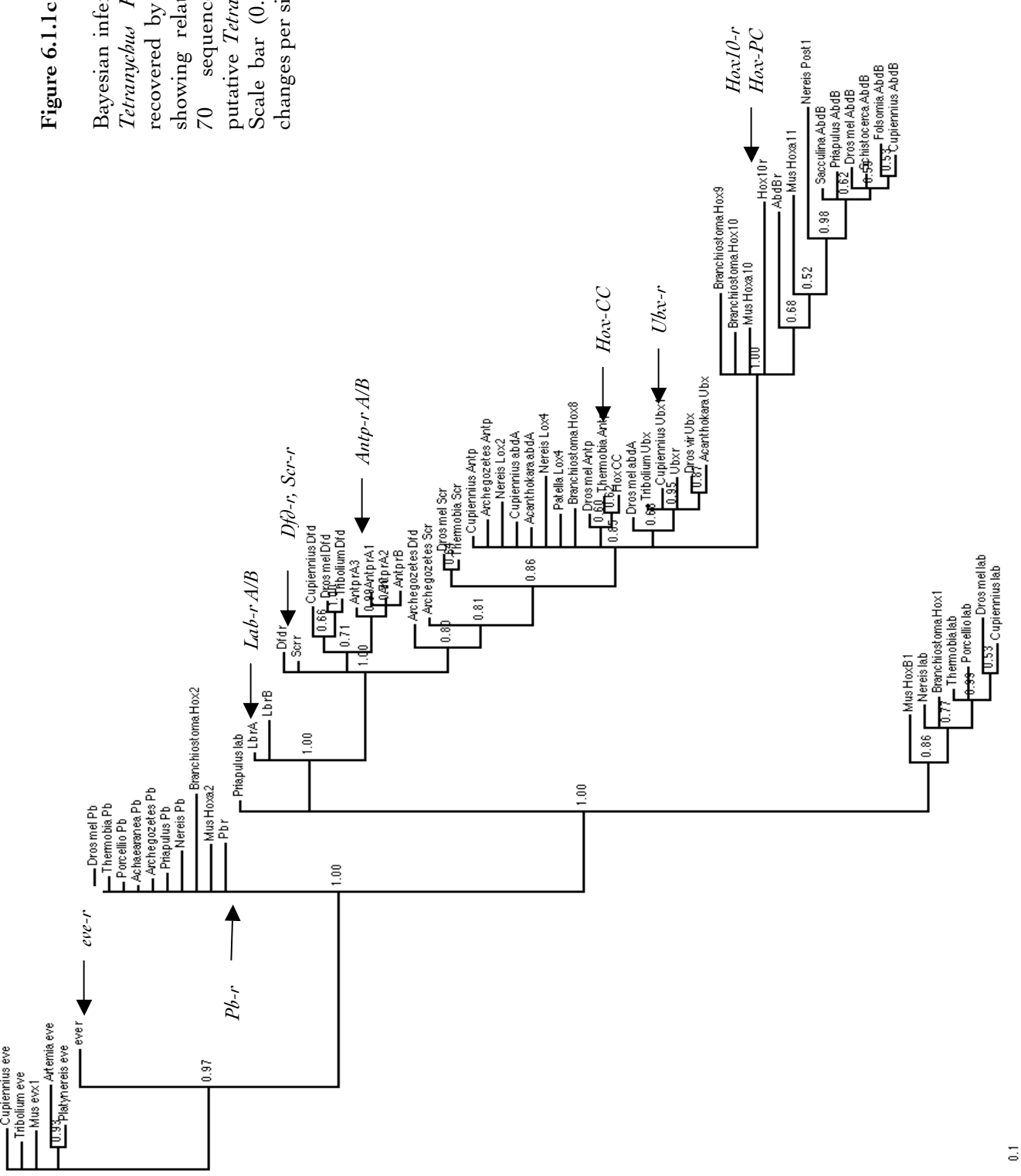
consensus tree topology (Figure 6.1.1c) receives high-maximal posterior probability at major nodes (between 0.86 and 1.00), supporting the affinity of *Tu-lab-r* to *labial/Hox1*, *Tu-pb-r* to *proboscipedia/Hox2*, *Tu-Ubx-r* to *Ultrabithorax/Hox7* and placing both *Tu-Hox-PC* and *Tu-Hox10-r* to an *AbdB/Hox9-10* clade. *Tetranychus Dfd-r* and *Scr-r* fall within an incompletely resolved [*Dfd+Scr*] clade, in which *Tu-Antp-r* genes somewhat surprisingly also appear. The *Hox* central class (*Hox-CC*) gene receives 0.85 posterior probability support as branching within a broad [*Antp+Lox2/4+Ubx+abdB*] clade (Figure 6.1.1c).

Although most of the putative *Tetranychus urticae* *Hox* genes can be assigned to a specific *Hox* cluster paralogy group on the basis of Bayesian analysis, the precise identities of *Tu-Antp-r* and *Tu-Hox-CC* remain uncertain beyond their being likely central class *Hox* orthologs. Some difficulty in conclusive *Hox* orthology assignment is not surprising given the sequence data analysed: the preliminary homeobox fragments are both short (180b) and highly conserved, predictably limiting phylogenetic signal. For these reasons, all orthology designations are conditional upon confirmation with extended sequence and all putative *Tetranychus* *Hox* genes are therefore named provisionally as *Hox*-‘related’ (*Hox-r*).

6.1.2 *Tu-Ubx* sequence extension and analysis

i) Inverse PCR results

I digested *Tetranychus* genomic DNA with restriction enzymes (BamHI, EcoRV, NcoI, RsaI, XhoI) that would not cut within the known *Tu-Ubx* homeobox region, and DNA fragments were then circularised and concentrated according to protocols based on those in Averof & Akam (1993) and described in Chapter VIII, section 8.2.3(Averof and Akam 1993). I designed three pairs of unique, gene-specific primers (in Appendix 2.7) for amplification outward from the *Ubx* sequence obtained *via* degenerate PCR into novel up/down-stream regions. PCR reagents and conditions were as detailed in Materials & Methods, 2nd PCR programs being either standard or including touchdown cycles, and annealing at either 48°C or 50°C (see Chapter VIII, section 8.2.1). DNA fragments between 350 - 650bp were amplified, sequenced and converted into genomic orientation to give a new total of 876bp *Tu-Ubx* (see Figure 6.1.2a,i). Novel sequence includes 315bp data 5’ of the homeobox and 470bp in the 3’ direction, within which is the *Ubx*-specific motif QAIKELNAQSK as defined in de Rosa et al. (1999) (de Rosa et al. 1999).



(i) *Tu-Ubx*:

```
1 31 61
GGT TAC ACA TTG AAT GGG TAA CCT TCC AAA AGC CAT TAG CTT ATT ACA ACA AAA AAT TGC GTC CAA AGA GTC AAA GTG TTG AAA GAT
91 121 151/2
GGA AAA GTA TCG AAA GAA GAA AAA AGT TTC CAA AAA ATA CAT TTG CCT AAT CTG ATG TTT TCT TCT CCT TTC TCT TCT TCT ACT CCT
241/42
181/22 211/32
TTT CTT CTC TTC TTC TCT [CTC TCT] CTT TCA ATG GTG CTT TTT TCT TTC GCA TCT CCT GTC GGT ATG ACA TGT TTC CCT GGC GGC AAT
phe leu leu phe phe phe ser leu ser leu ser met val leu phe ser phe ala ser pro val gly met thr cys phe pro gly gly asn
271/52 301/62 331/72
GGA CTG CGT CGT AGA GGT CGT CAG ACC TAT ACG CGC TAT CAG ACT CTT GAG CTA GAA AAA GAG TTT CAC ACA AGT CAA TAT TTG ACT CGT
gly leu arg arg gly arg gln thr tyr thr arg tyr gln thr leu glu leu glu lys glu phe his thr ser gln tyr leu thr arg
361/82 391/92 421/102
CGA CGT CGC ATT GAA CTG GCT CAC GCT CTT TGC CTA AGT GAA CGT CAG ATT AAA ATC TGG TTC CAA AAT CGT CGG ATG AAA CAG AAG AAG
arg arg arg ile glu leu ala his ala leu cys leu ser glu arg gln ile lys ile trp phe gln asn arg met lys gln lys lys
451/112 481/122 511/132
GAA ATT CAG GCA ATC AAA GAG CTC AAT GCA AGT AAA TCT TCT CAG TCA TCA TCA TCA TCA TCA TCC AAC AAT CAA TCA AGC AAC
glu ile gln ala ile lys glu leu asn ala gln ser lys ser ser gln ser thr ser ser ser asn asn gln ser ser asn
541/142 571/152 601/162
ACC AAC TCA TCG CGT AAC CAT CGA AGT GGT TCA GGT ATA ACT GAA CAC TCA AAT GGC TTG AAA CAC GAA TAA CAT AAA ATA ATG GAA AGA
thr asn ser ser arg asn his arg ser gly ile thr glu his ser asn gly leu lys his glu TAA -X-
631 661 691
CCA AAA GAC CAT ATA TAA AGA CCA TCG ACA GGT GAA AAT ACT CGA AAT ACT TTA TCA TCC TTT GCC AAA ACT GGA GGA AAA AAA TAG CGA
721 751 781
CAC AAG AAT CCA AGT CAT CAT ATT TTC ATC ATC AAT GGA CAC AAA TCA TCC ATC GCC ATT CAT CAC CAT CAT CAC CAT TGG ATT
811 841 871
AAG GGA AAA TCA AGG CCA ACA CTG ACA CAC AAA TTA CAG ACC AAC GCA AAC TGT AAC GTC AGA
```

(ii)

```
1 11 21 31 41 51 61 71
Ubx155 MFSFSPFSFT PFLIFFFSLS LSMVLFSSFA SPVGMTCFPG GNGLRRRGRQ TYTRYQTLEL EKEFHTSQYL TRRRRIELAH
Ubx153 MFSFSPFSFT PFLIFFFS-- LSMVLFSSFA SPVGMTCFPG GNGLRRRGRQ TYTRYQTLEL EKEFHTSQYL TRRRRIELAH
81 91 101 111 121 131 141 151
Ubx155 ALCLSERQIK IWFQNRMRKQ KKEIQAIKEL NAQSKSSQST SSSSSSSNNQS SNTNSSRNHR SGSGITEHSN GLKHE/
Ubx153 ALCLSERQIK IWFQNRMRKQ KKEIQAIKEL NAQSKSSQST SSSSSSSNNQS SNTNSSRNHR SGSGITEHSN GLKHE/
```

Figure 6.1.2a Nucleotide and amino acid translation data for *Tu-Ubx* sequence recovered by inverse PCR. **Bold:** homeodomain. Underlined: previous sequence obtained by degenerate PCR. **Purple:** Ubx signature peptide motif; **green:** 1st start codon; **red:** stop codon. **(i)** 876bp *Tu-Ubx*, including complete CDS, 5' and 3' UTR for 'Ubx¹⁵⁵' allele. Code missing in 'Ubx153' allele =CTC TCT. **(ii)** Ubx¹⁵⁵ allele vs. Ubx¹⁵³: grey = additional LS repeat in Ubx¹⁵⁵ prior to methionine residue #2.

ii) Multiple *Ubx* alleles?

Among the many clones obtained, I detected two forms of *Ubx* differing by two amino acids prior to methionine #2. I named the alleles *Ubx*¹⁵⁵ and *Ubx*¹⁵³ after their distinct polypeptide chain sizes: isoform *Ubx*¹⁵³ lacks a leucine-serine motif at residue 22-23, present in *Ubx*¹⁵⁵ as the second of two such repeats (Figure 6.1.2a). The repetitive nature of the sequence insertion (3x CT) suggests that it could represent a sequencing error. Alternatively, the SLSLS vs. S--LS coding may be of functional significance, providing physical elongation, higher polarity (serine –OH group) or an active site for a specific, unknown co-factor. The two transcripts could hypothetically be differentially regulated by i) some kind of alternative splicing if derived from the same gene, or ii) *via* distinct promoters/enhancers, activating duplicated genes in a time- or tissue-specific manner (Alberts et al. 1994; Davidson and Erwin 2006; Lewin 2002). The latter option (ii) seems very unlikely given the identical nucleotide sequence in the remaining coding regions of *Ubx*¹⁵⁵ and *Ubx*¹⁵³. However, several cases of independent *Hox* gene duplication within single chelicerate *Hox* clusters have been reported, such as in the Xiphosuran *Limulus polyphemus*, spiders *Cupiennius salei* (*Ubx*), *Steatoda triangulosa* (*Antp*) and *Achaearanea tepidariorum* (*pb*, *Dfd*), and arguably in *Tetranychus* (central-class *Hox* as well – c.f. Figure 6.1.1a given degenerate PCR results (Cartwright et al. 1993; Cook et al. 2001; Damen et al. 1998).

iii) Phylogenetic analysis of *Tu-Ubx*

A 195bp stretch of *Tu-Ubx* was aligned with sequence coding for the homeodomain and immediately flanking residues for *Ubx*, *labial*, *Scr* and *abðA* orthologs in a range of protostomes and deuterostomes. See Figure 6.1.2b for corresponding multiple protein sequence alignment, including relevant gene-specific signature residues and another *Tetranychus Hox* gene, *Tu-Scr*, that is mentioned further in **Appendix 3**. Bayesian inference analysis of the aligned nucleotide dataset gave a consensus tree (Figure 6.1.2c) in which *Tu-Ubx* branches with 0.99 posterior probability support in a monophyletic *Ubx* orthology clade, separate from distinct clades of other paralogy group representatives. This very well supported topology clearly indicates that the proposed *Tu-Ubx* gene is indeed a true spider mite *Ultrabithorax* ortholog.

	Helix1-----	Helix2----	Helix3-----	
Dm lab	NSGRTNFTNKQLTELEKEFHFNRYLT	RARRIEIANTLQLNETQVKIWFQNR	RMKQKRVKEGLIP	
Tc lab	NTGRTNFTNKQLTELEKEFHFNKYL	RARRIEIASALQLNETQVKIWFQNR	RMKQKRMKEGLIP	
Td lab	NAGRTNFTNKQLTELEKEFHFNKYL	RARRIEIASALQLNETQVKIWFQNR	-----	
Pc lab	GTGRTNFTTKQLTELEKEFHFNKYL	RARRIEIASALQLNETQVKIWFQNR	-----	
Lf lab	-----	HNKYLTRARRIEIATLQLNETQVKIWFQNR	RMKQKRLKEGSIL	
Cs lab	GSGRTNFTTKQLTELEKEFHFNKYL	RARRIEIATLQLNETQVKIWFQNR	RMKQKRMKEGLLV	
St lab	-----	HNKYLTRARRIEIASLGLNETQVKI	-----	
Pria lab	-----	HNKYLTRARRIEIAASLGLNETQVKI	-----	
Nv lab	---RTNFTNKQLTELEKEFHFNKYL	RARRIEIAAALGLNETQVKIWFQNR	RMKQKRMKETNVS	
Mm HoxB1	GGLRTNFTTKQLTELEKEGFHNKYL	SARRVEIAATLNETQVKIWFQNR	RMKQKREEREGGRM	
Bf Hox1	---RTNFTTKQLTELEKEFHFNKYL	SARRVEIAAALNLETQVKIWFQNR	RMKQKREKENGFS	
Ak Scr	KRQRTSYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQIKIWFQNR	MKWKKEHKMSSST	
Dm Scr	KRQRTSYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQIKIWFQNR	MKWKKEHKMASMN	
Tc Scr	KRQRTSYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQIKIWFQNR	MKWKKEHKMASMN	
Td Scr	KRQRTSYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQIKI	-----	
Af Scr	KRQRTSYTRYQTELEKEFHFNRYLT	RRRIEIAHTLCLSERQIKIWFQNR	MKWKKEHKIATMN	
Sc Scr	KRQRTSYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQVKIWFQNR	MKWKKEHKMANMG	
Er Scr	----TSYTRYQTELEKEFHFNRYLT	RRRIEIAHSLCLSERQIKIWFQNR	MKWKKEHKIPSVN	
Al Scr	KRQRTSYTRYQTELEKEFHFNRYLT	RRRIEIAHSLCLSERQIKIWFQNR	MKWKKEH-----	
At Scr	-----	HNHYLTRRRRIEIAHALCLTERQIKI	-----	
Tu Scr	-----	TRYQTELEKEFHFNRYLT	RRRIEIAHSLCLSERQIKIWFQNR	MKWKKEIKFQHGS
Pria HB1	-----	EFHFNRYLT	RRRIEIANVLRTERQIKIWFQNR	MKWKKEHRLPKPV
Nv Scr	KRTRTSYTRHQTELEKEFHFNRYLT	RRRIEIAHALNLTERQIKIWFQNR	MKWKKEHKLGHLA	
Hr Lox20	KRTRTSYTRHQTELEKEFHFNRYLT	SRRRIEIAHVNLNLSERQIKIWFQNR	MKWKKDHPNHPGS	
Mm HoxA5	KRARTAYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLSERQIKIWFQNR	MKWKDNLKLSMS	
Bf Hox5	KRTRTAYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQIKIWFQNR	MKWKKENKLKLSLS	
Dm Ubx	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Am Ubx	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHSLCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Tc Ubx	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Fc Ubx	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Af Ubx	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHSLCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Sc Ubx	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHQLCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Er Ubx	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Cs Ubx1	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEIQAIKEL	
Cs Ubx2	-RGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHSLCLTERQIKIWFQNR	MKLKKEAQAIKEL	
Tu Ubx	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLSERQIKIWFQNR	MKQKKEIQAIKEL	
Ak Ubx	KRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEIMOTIKDL	
Pria UbdA	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEISQALCLTERQIKIWFQNR	MKLKKEIQAALKEM	
Pv Lox2	-----	FKFNRYLT	RRRIEISHLCLTERQIKIWFQNR	MKEKKELQAIKEL
Nv Lox2	-RGRQTYTRYQTELEKEFHFNRYLT	RKRRIEISHLCLTERQIKIWFQNR	MKEKKELQAIKDL	
Pv Lox4	-----	FQFNHYLTRKRRIEIAHTLCLTERQIKIWFQNR	MKMRERQAIKDI	
Nv Lox4	ERGRQTYTRYQTELEKEFHFNHYLT	RKRRIEIAHVLCLTEHQIKIWFQNR	-----	
Sp Hox1	KRCRQTYTRYQTELEKEFHFNRYLT	RRRIEISHLGLTERQIKIWFQNR	MKYKKEKSNKEEG	
Hs HoxA7	KRGRQTYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQIKIWFQNR	MKWKKEHKDEGPT	
Gg HoxA7	KRGRQTYTRYQTELEKEFHFNRYLT	RRRIEIAHALCLTERQIKIWFQNR	MKWKKEHKKESSS	
Dm AbdA	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEELRAVKEI	
Tc AbdA	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEELRAVKEI	
Sa AbdA	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEELRAVKEI	
Af AbdA	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEELRAVKEI	
Cs AbdA	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHALCLTERQIKIWFQNR	MKLKKEEMRAVKEI	
Ak AbdA	RRGRQTYTRYQTELEKEFHFNHYLT	RRRIEIAHVLCLTEQIKIWFQNR	MKLKKEELRAVKEI	
Mm HoxB8	RRGRQTYTRYQTELEKEFLFNPYLT	RKRRIEIVSHALGLTERQVKIWFQNR	MKWKKEENNKDKFP	
Hs HoxD8	RRGRQTYTRYQTELEKEFLFNPYLT	RKRRIEIVSHALGLTERQVKIWFQNR	MKWKKEENNKDKFP	
Bf Hox8	RRGRQTYTRYQTELEKEFHFNKYL	RRRIEIAHALGLTERQIKIWFQNR	MKLKKEEAAMLCP	

Figure 6.1.2b Multiple protein sequence alignment (55x 65 amino acids) to compare *Tu*-Ubx and *Tu*-Scr residues alignable with orthologs from a range of protostome and deuterostome taxa. Bold: *Tetranynchus* Hox proteins. Grey: homeodomain sequence. Red font: α -helices within HD. Signature residues/motifs diagnosing specific Hox proteins: Light blue – Lb; Purple – Scr; Green – Ubx; Yellow – UbdA; Blue – AbdA.

Species abbreviations: Dm – *Drosophila melanogaster*; Tc – *Tribolium castaneum*; Td – *Thermobia domestica*; Pc – *Procambarus clarkii*; Lf – *Lithobius forficatus*; Cs – *Cupiennius salei*; St – *Steatoda triangulosa*; Pria – *Priapulus caudatus*; Nv – *Nereis virens*; Mm – *Mus musculus*; Bf – *Branchiostoma floridae*; Ak – *Akantorhynchus kaputensis*; Af – *Artemia franciscana*; Sc – *Sacculina carcini*; Er – *Etmopterus rubripes*; Al – *Archegozetes longistosus*; At – *Achaeanea tepidariorum*; Hr – *Halocynthia roretzi*; Am – *Apic mellifera*; Fc – *Folsomia candida*; Pv – *Patella vulgata*; Sp – *Strongylocentrotus purpuratus*; Hs – *Homo sapiens*; Gg – *Gallus gallus*; Sa – *Schistocerca americana*.

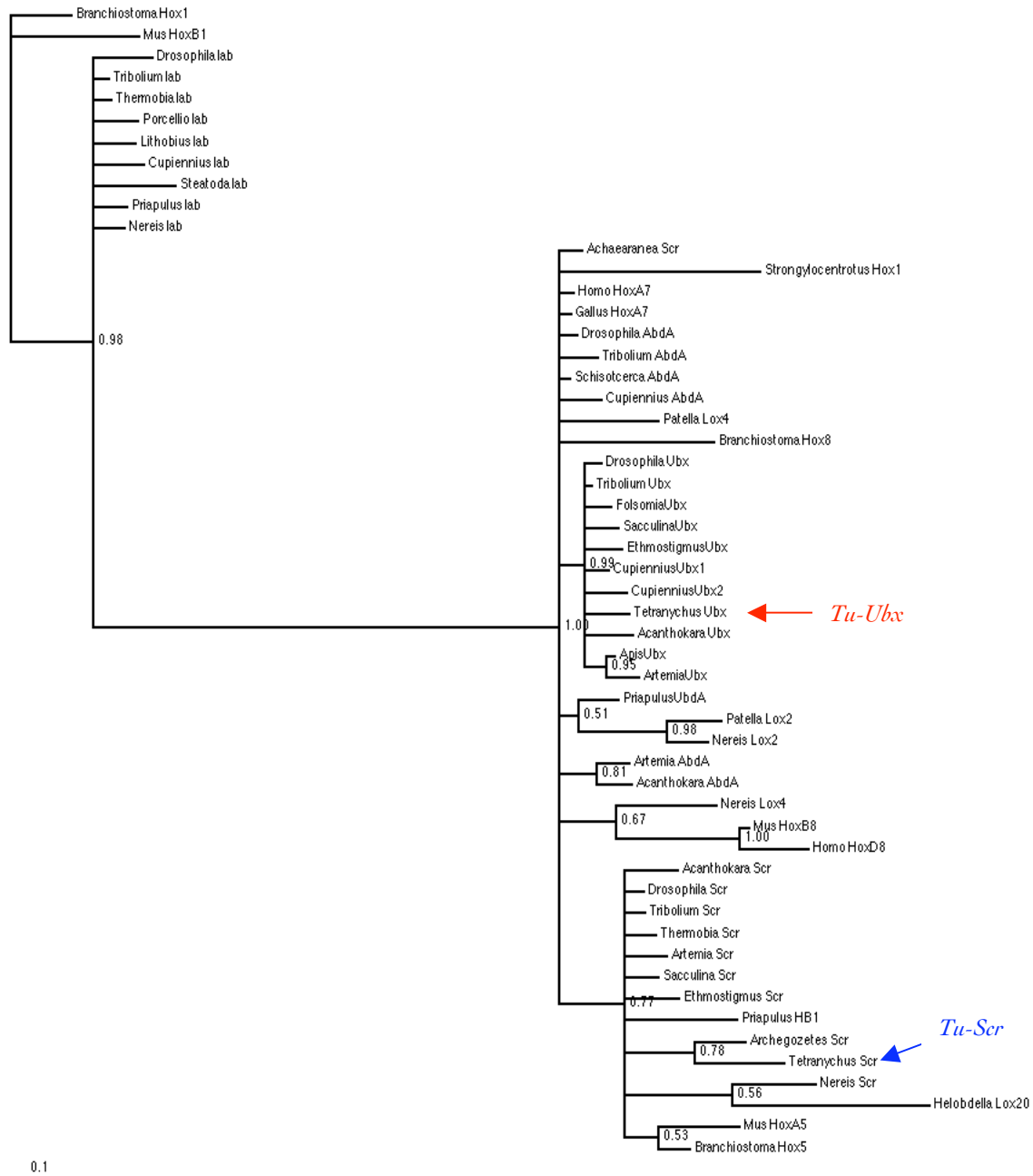


Figure 6.1.2c Bayesian consensus tree for analysis of 195bp *Tu-Ubx* (red) and *Tu-Scr* (blue) in relation to other metazoan *Hox* genes from *labial/Hox1*, *Scr/Hox5*, *Ubx/Hox7* and *abdA/Hox8* paralogy groups. 55 sequences, 3 from *Tetranychus*.

iv) Assessment of Ubx C-terminal domain function

Model based on Drosophila and Artemia Ubx

It has been proposed that differences in Ubx C-terminal protein residues between *Akanthokara* (onychophora), *Artemia* (crustacean) and *Drosophila* (insect) explain the limbed or limbless status of segments, associated with Ubx' ability to repress the *Dll-504* early enhancer or not (Galant and Carroll 2002; Hsia and McGinnis 2003; Hughes and Kaufman 2002b; Levine 2002; Pavlopoulos and Averof 2002; Ronshaugen et al. 2002; Vervoort 2002). *Artemia* Ubx contains a serine-threonine rich C' domain, providing consensus sites for kinase-mediated phosphorylation, for example by CKII (S/T-X-X-D/E), GSK3 (S/T-X-X-S/T) or MAP kinase (S/T-P)(Fiol et al. 1987; Jaffe et al. 1997). S/T kinase phosphorylation of *Artemia* Ubx is proposed to inhibit a latent repression element elsewhere in the protein, thus preventing *Dll* repression and allowing the appendages to develop on the crustacean thorax. In *Drosophila* Ubx, S/T residues are presumed to have been gradually replaced by a glutamine-alanine motif (QAQAQK) and poly-alanine domain, releasing potent *Dll* repression throughout the consequently limbless fly abdomen (Galant and Carroll 2002; Hughes and Kaufman 2002b; Ronshaugen et al. 2002). I compared Ubx protein C terminal regions from a wide range of arthropods (Figure 6.1.2d), including *Tetranychus urticae* and outgroup onychophoran taxon *Akanthokara*, in order to assess the validity of this scenario.

Hexapod, crustacean and myriapod Ubx proteins

It is clear from the multiple sequence alignment (Figure 6.1.2d) that the QAQAQK motif adjacent to the UbdA motif is ancestral to the arthropods, rather than being an insect innovation as suggested by previous authors, and hence cannot be implicated in *Dll* repression/de-repression alone. Within **Hexapoda**, a poly-A stretch of 9-10 alanines correlates with Ubx ability to repress *Dll* via DNA binding as a Hox/Exd/Hth complex, whereas in *Tribolium* the poly-A stretch is interrupted at site 8 by a valine residue, perhaps partially¹⁹ explaining lack of *Dll* repression by *Tc*-Ubx, with its role in A1 segment appendage modification instead (Bennet et al. 1999; Lewis et al. 2000; Palopoli and Patel 1998). Within **Crustacea**, the QAQAQK motif is variously lost or retained, and numerous sites of S/T kinase activity (CKII, GSK3, MAPK-type) and limited C' alanine residues are present. *Ubx* and *Dll* are co-expressed in the limbed crustacean thorax, and Ubx C' terminal domain structure supports selection for S/T phosphorylation to inhibit *Dll* repression, but indicates that the QAQAQK motif and the type of phosphorylation achieved, are selectively neutral in this process (Abzhanov and Kaufman 2000a; Averof and Akam 1995; Hsia and McGinnis 2003; Mouchel-Vielh et al. 1998; Ronshaugen et

¹⁹ *Tribolium* Ubx provides full phenotypic rescue when expressed ectopically in *Drosophila* Ubx mutants, indicating that different *Dll*-repression function between *Tc*-Ubx and *Dm*-Ubx largely depends upon changes *in cis*, in the responsiveness of particular target genes (i.e *Dll*) to otherwise equivalent or interchangeable Hox proteins (Galant et al. 2002).

		1	20	40	60	
Dm-Ubx	Hexapoda	MAHALCLTERQ	I	KI	WFQ	NR
Dv-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Tc-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Jc-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Am-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Fc-Ubx	Collembola	MAHALCLTERQ	I	KI	WFQ	NR
Af-Ubx	Crustacea	MAHSLCLTERQ	I	KI	WFQ	NR
Sc-Ubx		MAHQLCLTERQ	I	KI	WFQ	NR
Ps-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Tc2-Ubx		MAHQLCLTERQ	I	KI	WFQ	NR
Er-Ubx	Myriapoda	MAHALCLTERQ	I	KI	WFQ	NR
Ag-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Gm-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
La-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Sm-Ubx		MAHALCLTERQ	I	KI	WFQ	NR
Cs-Ubx1	Chelicerata	MAHALCLTERQ	I	KI	WFQ	NR
Cs-Ubx2		MAHSLCLTERQ	I	KI	WFQ	NR
Tu-Ubx		LAHALCLSERQ	I	KI	WFQ	NR
Ak-Ubx	Onychophora	MAHALCLTERQ	I	KI	WFQ	NR

Figure 6.1.2d Multiple alignment of 14 arthropod/onychophoran partial Ubx proteins, including 27 amino acids of 3' homeodomain (grey), Ubx signature peptide (purple) and final C terminal residues. Blue: shared identity to *Drosophila/Junonia* QAQAK+10x/12x alanine motif. Red font: all sites of putative ser/thr kinase mediated phosphorylation. Yellow: MAP kinase phosphorylation consensus site (thr-pro/ser-pro). Green: casein kinase II (CKII) consensus phosphorylation site (ser/thr-X-X-asp/glu). Boxed: GSK-III consensus phosphorylation sites, overlapping where multiple in succession (ser/thr-X-X-ser/thr).

Species: Dm = *Drosophila melanogaster*; Dv = *Drosophila virilis*; Tc = *Tribolium castaneum*; Jc = *Junonia coenia*; Am = *Apis mellifera*; Fc = *Folsomia candida*; Af = *Artemia franciscana*; Sc = *Sacculina carini*; Ps = *Porcellio scaber*; Tc2 = *Trigriopus californicus*; Er = *Ethmostigmus rubripes*; Ag = *Archispirostreptus gigas*; Gm = *Glomeris marginata*; La = *Lithobius atkinsoni*; Sm = *Strigamia maritima*; Cs = *Cupiennius salei*; Tu = *Tetranynchus urticae*; Ak = *Akanthobkara kaputensis*.

al. 2002). Alternatively, in the absence of functional data for species other than *Artemia*, unknown independent mechanisms could confer activation status to Ubx or disrupt its repression capability. In **Myriapoda**, where *Ubx* and *Dll* are also co-expressed throughout the trunk, the Ubx QAQAAK motif is partially diverged in sequence, and extensive GSK3 and MAPK consensus phosphorylation sites are evident (Figure 6.1.2d) (Brena et al. 2006; Grenier et al. 1997; Hughes and Kaufman 2002a; Hughes and Kaufman 2002b). This could also support a requirement for Ubx C' phosphorylation in order to overcome latent Dll repression, and if so, again highlights the likely selective insignificance of the QAQAAK domain. As in Crustacea, absence of functional data means alternative genetic mechanisms are also equally feasible.

Chelicerate Ubx proteins

In the chelicerates *Cupiennius* and *Tetranychus*, *Ubx* is expressed in the limbless opisthosoma, whereas early *Dll* expression is restricted to prosomal limb domains; a pattern reminiscent of Ubx-mediated *Dll* repression in the *Drosophila* abdomen (Damen et al. 1998). *Cupiennius* Ubx retains the ancestral QAQAAK motif, unlike *Tetranychus* Ubx, and both Ubx proteins contain an extended and enriched proportion of S/T consensus phosphorylation sites, primarily of GSK3-type (Figure 6.1.2d). If we assume that spider and mite Ubx do repress *Dll*, it casts doubt on the idea that S/T kinase-mediated disruption of the Ubx-*Dll* protein-DNA interaction is the only mechanism by which this is achieved. It would infer an independent mechanism that permits S/T kinase activity without disrupting Ubx-*Dll* interactions, distinct from the poly-A facilitated mechanism documented in higher insects. Other examples of independent molecular means of evolving the same phenotype ('pheno-genetic drift') include abdominal prolegs in Lepidoptera vs. Symphyta, and multiple, convergent instances of limb loss in reptiles and amphibians (Wray, 2007; Suzuki, 2001; Weiss, 2000; Kearney, 2004; Apesteguía, 2006; Coates, 2000; Caldwell, 1997; Cohn, 1999; Lee, 2000; Greene, 2000). However, if the exclusive nature of early chelicerate *Dll* and *Ubx* expression domains does not in fact reflect the ability of *Tu*-Ubx to repress *Dll*, the model suggested on the basis of *Artemia* and *Drosophila* Ubx data remains tenable. Furthermore, the possibility that protein expression may not reflect obvious or predicted functional interactions serves as a reminder that the phenotypic effect(s) of protein evolution depend ultimately on responsiveness of target gene *cis*-regulatory elements to change; Ubx protein may evolve to be a potential repressor, but if it is expressed without appropriate cellular (e.g. kinase) and target genetic (e.g. *cis*-regulatory) factors, this ability will go un-realised.

Limited conclusions...

Without functional data for chelicerate Ubx it is not possible to make conclusions about the functional significance of Ubx protein evolution in arthropods. However, the diversity in arthropod Ubx C-terminal sequences that is revealed in Figure 6.1.2d shows the utility of broad taxonomic

comparison when constructing evolutionary models, even if only to complicate or call into question those based on more limited data.

6.1.3 *Tu-Ubx* single-stranded RNA probe synthesis

Tu-Ubx plasmid preparation

Based on the ~0.9Kb *Tu-Ubx* genomic sequence determined by inverse PCR, I designed 8 gene-specific primers, enabling amplification and cloning of target *Tu-Ubx* gene fragments into pGEM®-T Easy vector for subsequent RNA probe synthesis (primer details in Appendix 2). A range of *Tu-Ubx* DNA fragments 511bp – 798bp in size were amplified by standard PCR on genomic DNA, encompassing the complete *Tu-Ubx* CDS plus minimal 5'/3' UTR material. See Figure 6.1.3a for a scaled schematic of *Tu-Ubx* transcript structure, corresponding gene-specific primer locations and amplification products. To confirm expected *Tu-Ubx* insert size and adequate DNA concentration for DIG-labelled ssRNA probe synthesis, I excised inserts by EcoRI digestion of purified plasmid DNA samples (Figure 6.1.3b,i-ii).

Tu-Ubx RNA probe synthesis and DIG-AP labelling verification

Depending on *Tu-Ubx* insert orientation in pGEM®-T Easy vector and whether sense or anti-sense RNA was to be synthesised, plasmids were linearised by ApaI/PstI restriction digest at one end of the insert, and RNA transcription mediated by Sp6/T7 respectively (see Figure 6.1.3b,i for specific conditions). Once synthesised, I tested *Ubx* probe RNA for effective DIG-labelling by hybridising serially diluted, membrane-bound samples with DIG-AP antibody. The results of these α -DIG-AP binding reactions were positive (Figure 6.1.3b,iii), showing declining phosphatase-dependent signal intensity with increasing labelled probe dilution: this confirms a theoretical readiness to detect target gene transcription *in situ*.

6.1.4 Absence of *Tu-abdA* from the *Tetranychus urticae* genome

i) Degenerate PCR with *Hox* homeodomain primers

A spider mite *abdominal-A* ortholog was not recovered during PCR screening of genomic DNA with general degenerate *Hox* homeobox primers (Hox-F1, Hox-R1 and R2; see **Appendix 2.7** for primer details). To confirm or refute the possible absence of *abdominal-A* that this negative result suggests, I

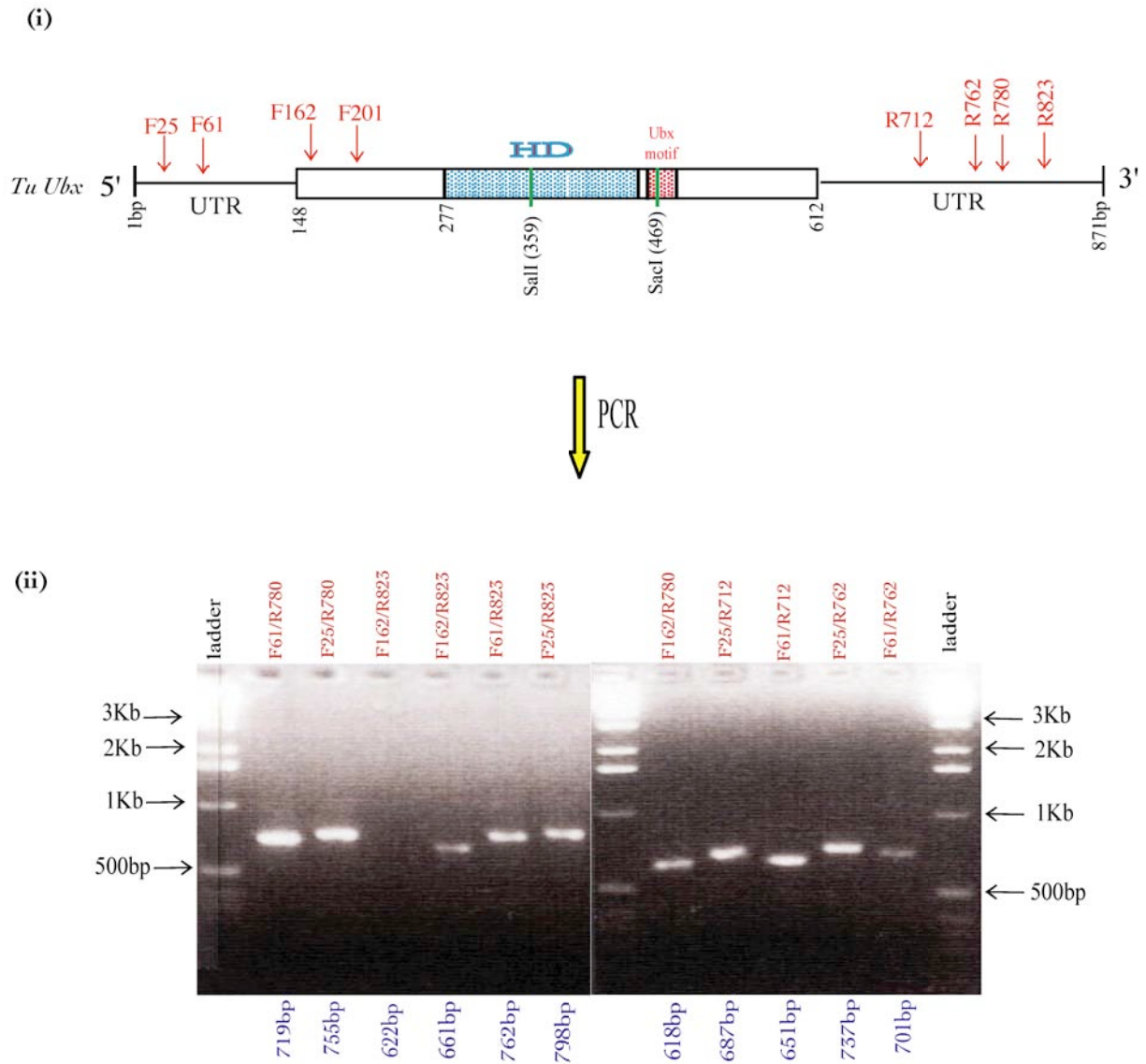


Figure 6.1.3a (i) *Tu-Ubx* gene structure for total sequence data obtained. 464bp continuous ORF including homeodomain (HD, blue stippling), Ubx signature motif (red stippling) and with enzyme cut sites marked. (ii) Gene specific primers (red arrows) were used to amplify Ubx fragments between 618bp and 798bp size by standard PCR, followed by purification and cloning of products for use in RNA probe synthesis.

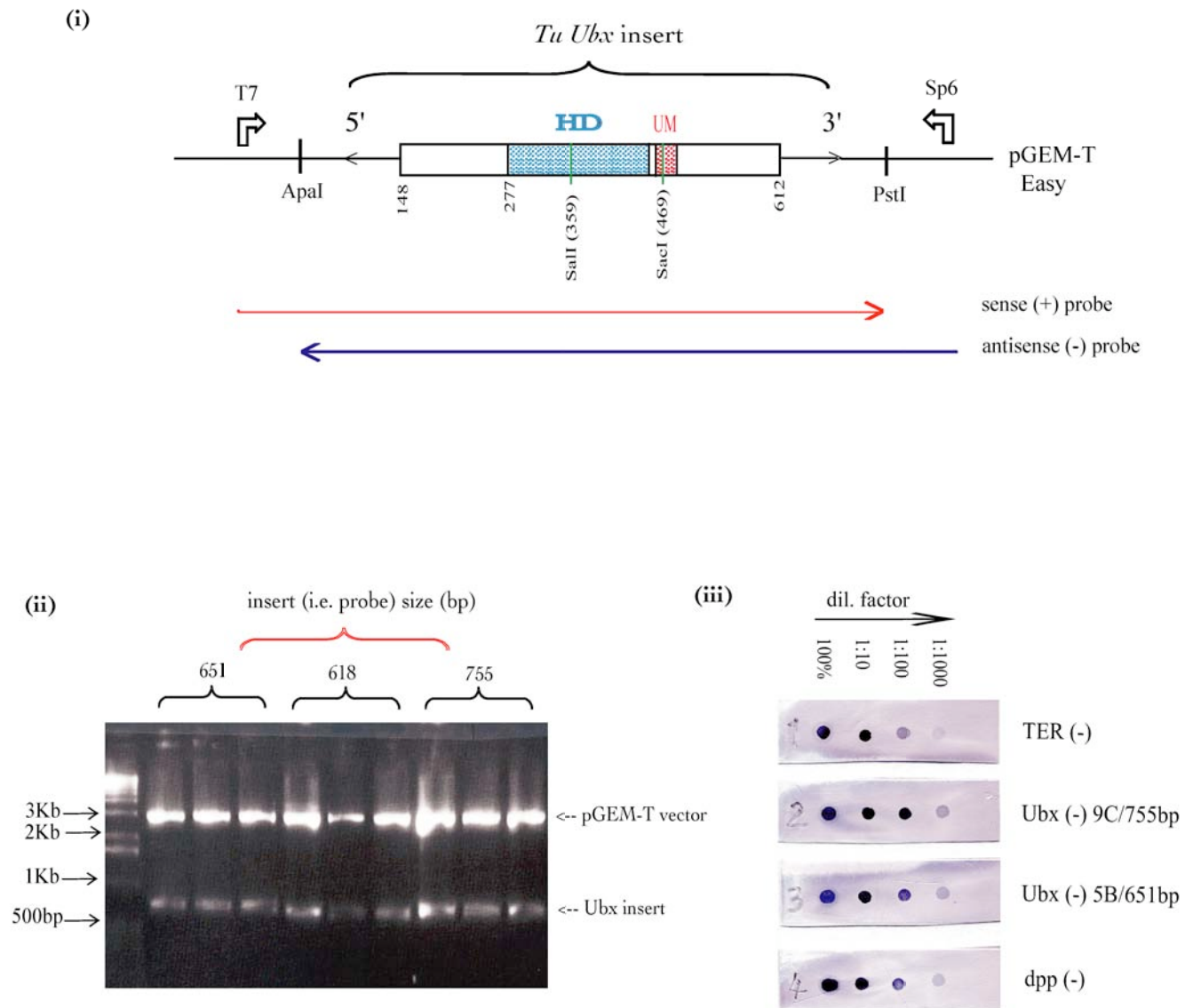


Figure 6.1.3b (i) Schematic (*Ubx* ORF to scale) of *Tu-Ubx* insert into pGEM-T Easy vector, with restriction enzymes and RNA polymerases for RNA probe synthesis from an insert in 5'-3' orientation. HD: homeodomain; UM: *Ubx* signature motif. (ii) Agarose gel run of *EcoRI* digestion, excising insert from vector to confirm correct insertion and DNA concentration prior to linearisation (with *ApaI*/*PstI*) and RNA probe transcription. (iii) Dot blot to test effective DIG-labelling of selected RNA probes; positive result of decreasing signal intensity with increasing diultion factor. Controls: TER and dpp (antisense).

carried out further genome screening with another set of primers, targeted to a conserved *Hox* motif 5' of the homeobox.

ii) Degenerate PCR with YPWM motif primers

All Hox proteins except those of the AbdB/Hox9 class contain a conserved hexapeptide motif (PRYPWM) upstream of the homeodomain, connected to it by a partially conserved 'linker' region of variable length (Merabet et al. 2003). Residues N' and C' to the YPWM are involved in stabilising Hox complexes on target *cis*-regulatory element DNA, and the hexapeptide motif itself is implicated in:

- (i) Mediating Hox association with Pbc/TALE (Three Amino acid Loop Extension) class cofactors (e.g. Exd, Hth), creating trimeric complexes with increased DNA binding affinity and specificity (Merabet et al. 2003; Shanmugam et al. 1997).
- (ii) AbdA regulatory output control over target genes, achieved by interactions between the YPWM and linker PFER motif in response to detecting cofactors for repression/activation (Galant et al. 2002; Gebelein et al. 2002; Grienemberger et al. 2003; Merabet et al. 2003).

Based on an alignment of *Drosophila melanogaster*, *Tribolium castaneum* and *Cupiennius salei* AbdA proteins (Figure 6.1.4,i), I designed several low-degeneracy primers to amplify forwards from the hexapeptide coding region (primer sequences in Appendix 2.7). The generic Hox-R1 reverse primer (5'-GGA ATT CAT ICK ICK RTT YTG RAA CCA IAT YT-3') was used in combination with hexapeptide-specific forward primers, in standard or nested amplification reactions on genomic or cDNA template. I tested PCR programs either ascending or descending towards an annealing temperature of 50°C or 52°C, aiming to find conditions that would allow the general *Hox* primer and more stringent YPWM primers to both bind effectively. Magnesium ion concentrations between 2.5mM – 3.5mM were tested, and amplification shown to be most effective at >3.0mM [Mg²⁺].

In all cases of successful amplification, PCR products of variable size (200bp - ~650bp) were amplified (Figure 6.1.4, ii-iv), indicative of non-specific, non-target primer binding. This was confirmed by cloning and sequencing results, which were either unidentifiable or pertained to arthropod 18S rRNA.

iii) Loss of *abd-A* linked with body plan reduction?

As degenerate PCR screens with both *Hox* homeobox and hexapeptide motif-targeted primers yielded no putative *Tu-abdominal-A* DNA fragments, it appears likely that the *abdominal-A* gene is

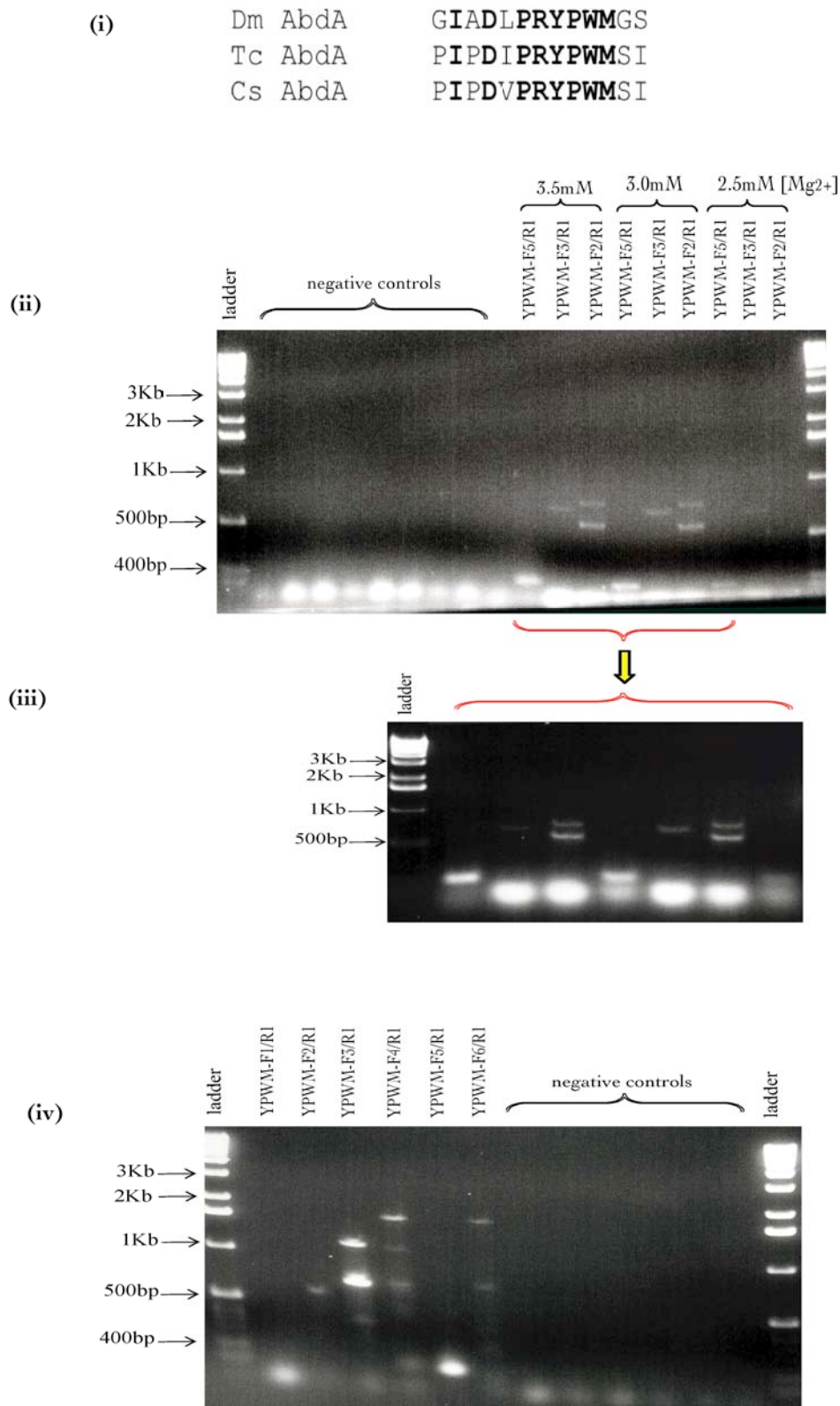


Figure 6.1.4 Degenerate PCR with Hox + YPWM motif primers to screen for abdA in *Tetranychus* genomic DNA. (i) Alignment of arthropod AbdA proteins - *Drosophila melanogaster* (Dm), *Tribolium castaneum* (Tc) and *Cupiennius salei* (Cs). (ii)&(iii) 'downdeg' PCR with YPWM Fwd primers + QNK Rev, at 2.5mM, 3.0mM and 3.5mM magnesium ionic concentration. (iii) 'updeg' PCR testing all YPWM Fwd primers + Rev QNK. The multiple products were 18S rRNA-derived, i.e. non-target artefacts.

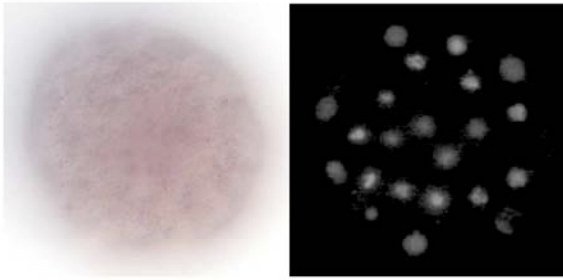
missing from the *Tetranychus urticae* genome. The loss of this posterior class gene, whilst retaining *Ultrabithorax* and *Abdominal-B*, may be related to extreme reduction of the opisthosoma in the spider mite body plan. In support of this idea, another mite *Archegozetes longesitosus* has a similarly reduced opisthosoma and apparent parallel loss of *abdominal-A* from its *Hox* cluster (Telford and Thomas 1998a; Thomas and Telford 1999). Also, loss of *abd-A* has been independently associated with the extremely reduced abdomen of the Rhizocephalid crustacean *Sacculina carcini* (Blin et al. 2003; Gibert et al. 2000; Mouchel-Vielh et al. 2002; Mouchel-Vielh et al. 1998; Quéinnec, 1999). If shown to be genuinely causal in secondary loss of posterior body segments, absence of *abd-A* in independent arthropod lineages (Rhizocephalida: Crustacea vs. Acarida: Chelicerata) represents convergence on the same genetic mechanism for a similar phenotypic outcome. Such a situation is directly opposed to the pheno-genetic plasticity described previously for *Ubx*, in which divergent genetic mechanisms may produce the same phenotype.

6.2 *Tu-Ubx* mRNA transcription during embryogenesis

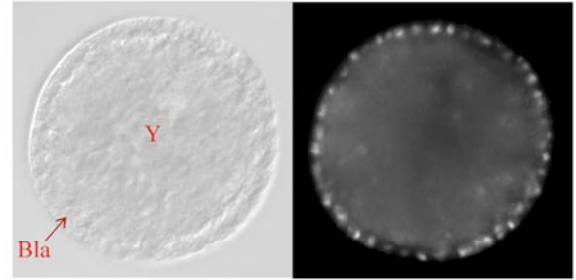
6.2.1 Lack of detectable *Tu-Ubx* transcripts: 0 – 17hr AEL

I did not detect *Tu-Ubx* transcription during cleavage, blastoderm stage, germdisc formation or early germband stages (Figure 6.2.1,i-iv). In *Drosophila*, *Ubx* mRNA is detected from late syncytial blastoderm stage onwards, associated with specifying posterior body (i.e. abdomen) identity within the long germband embryo (Akam and Martinez Arias 1985). A clear association between *Ubx* and posterior tagmatic specification has been demonstrated in species from the complete range of arthropod sub-phyla, gene activity correlated with emerging abdominal, pleon or trunk segments in insects, crustaceans, myriapods and chelicerates (reviewed in Cornec, 2006; Angelini, 2005; Hughes, 2002)(Abzhanov et al. 1999; Angelini and Kaufman 2005b; Angelini et al. 2005; Averof and Akam 1993; Averof and Akam 1995; Averof and Patel 1997; Bennet et al. 1999; Blin et al. 2003; Casares et al. 1996; Cornec and Gilles 2006; Damen et al. 1998; Deutsch and Mouchel-Vielh 2003; Galant and Carroll 2002; Grenier et al. 1997; Hughes and Kaufman 2001; Hughes and Kaufman 2002b; Kelsh et al. 1994; Lewis et al. 2000; Mouchel-Vielh et al. 2002; Palopoli and Patel 1998; Popadic and Nagy 2001; Ronshaugen et al. 2002; Suzuki and Palopoli 2001; Ueno et al. 1992; Weatherbee and Carroll

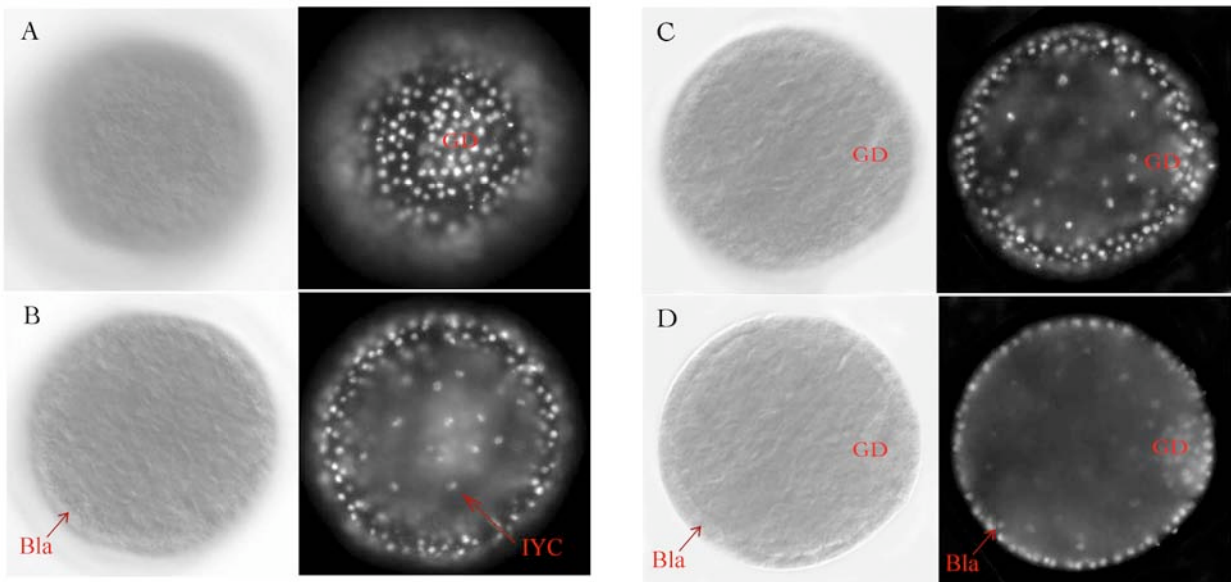
(i)



(ii)



(iii)



(iv)

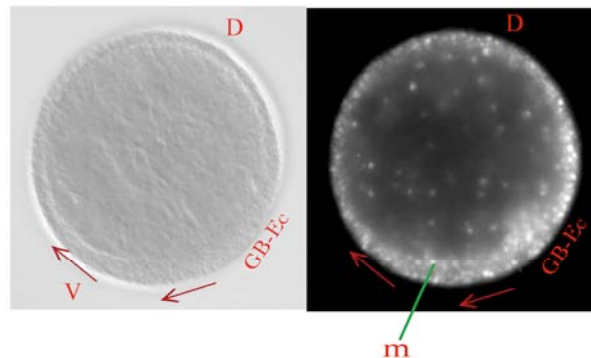


Figure 6.2.1 *Ubx* *in situ* hybridisation in blastoderm to early germband stage embryos: no transcripts are detected. (i) 64-cells. (ii) Blastoderm stage. (iii) Germdisc stage: A-B sections onto GD, C lateral, D sagittal sections. (iv) Early germband, sagittal.

D - dorsal; V - ventral; A - anterior; P - posterior; Bla - blastomere cell layer; Y - yolk; GD - germ disc; IYC - inner yolk cell; GB-Ec - germband ectoderm; m - mesoderm.

1999; Zhang et al. 2005). The *Tetranychus* germband grows from a posterior germdisc proliferation zone, with prosomal segments appearing first (Anderson 1973). Hence it is not surprising that *Ubx* is not expressed in the early spider mite embryo, as genetic delineation of posterior segment identity is presumably not yet active.

6.2.2 Early *Ubx* expression

i) Mid to late germband stage: 19-22hr AEL

I first detected *Tu-Ubx* transcripts at mid-late germband stage (~19-22hrs AEL), in a small, approximately semi-lunar domain localised most strongly to ectodermal cells of the presumptive anterior opisthosoma. The anterior boundary appears straight, forming a transverse division between emerging prosomal and opisthosomal territories. *Ubx* expression tapers slightly and fades posteriorly, presumably reflecting opisthosomal morphology at this time-point (Figure 6.2.2,i-ii).

ii) Late germband to early limb bud stage: 22-29hr AEL

Tu-Ubx mRNA transcripts remain localised in an opisthosomal ectoderm-restricted domain from late germband stage onwards, in which all prosomal segments are delineated with shallow grooves. A linear anterior border persists, *Ubx* activity tapers in line with morphology, and is absent from the most posterior opisthosomal or telson region (Figure 6.2.2,iii-iv). Double *in situ* hybridisation for *Tu-Ubx* and *Tu-Dll* gene activity confirms a gap between *Dll*-positive cells of the L3 limb primordium and the anterior boundary of *Tu-Ubx* expression (Figure 6.2.2,v). This indicates that *Tu-Ubx* transcription is absent from the most anterior segment(s) of the opisthosoma, in accord with a proposed ancestral anterior *Ubx* boundary in the Op2 segment of chelicerates (Damen et al. 1998; Popadic and Nagy 2001). Unfortunately, the extreme secondary reduction of the spider mite opisthosoma, and morphological fusion of segments means that I cannot determine precisely which segment the anterior *Tu-Ubx* boundary marks, and hence cannot easily discern spatial/temporal boundary modulation, as has been elegantly described for *Limulus* (Xiphosuran), *Paruroctonus* (scorpion) and several spiders (Abzhanov et al. 1999; Popadic and Nagy 2001). It would be ideal with respect to confirming segment boundaries to detect for *Tu-Ubx* mRNA in parallel with parasegment genes such as *engrailed*, although this is not possible at present (c.f. Chapter IV).

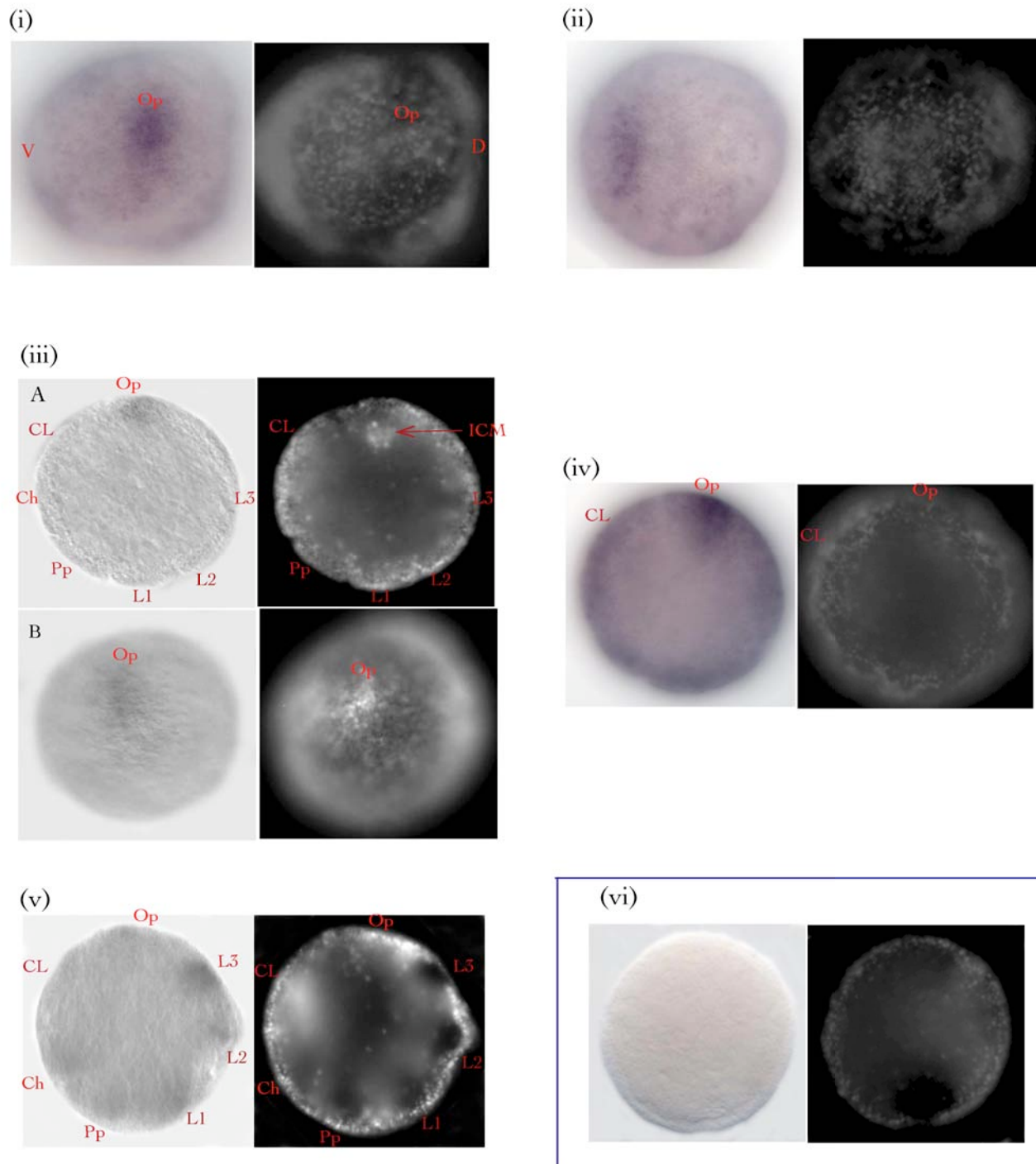


Figure 6.2.2 *Ubx* *in situ* mRNA transcription pattern from mid germband to earliest limb bud stage: Transcripts localise to a posterior/opsthosomal ectoderm domain, overlying the inner cell mass. The anterior border is transverse and approximately linear, and expression tapers towards the posterior, reflecting morphology of the reduced opisthosoma. (i)-(ii) *Ubx* antisense probe: mid germband stage. (iii)-(iv) *Ubx* antisense at late germband stage. (v) *Ubx* + *Dll* antisense double mRNA *in situ*: late GB-early limb bud stage. (vi) Negative control: sense probe.

D - dorsal, V - ventral, CL - cephalic lobe, Ch - chelicera, Pp - pedipalp, L1 - 1st walking leg, ICM - inner/endodermal cell mass, Op - opisthosoma

6.2.3 Later *Ubx* expression

i) *Tu-Ubx* during ventral (ridge) closure: 30-45hr AEL

While the ventral median ridge closes and prosomal limbs are elongating, I continue to detect *Tu-Ubx* mRNA transcripts in a domain restricted to opisthosomal ectoderm that appears weaker at the ventral midline (Figure 6.2.3,i-ii). In some ventral closure stage embryos *Ubx* activity is absent from terminal segment(s) and low at the posterior midline of the transcription domain, save for a strong terminal midline spot (Figure 6.2.3,ii). Lower apparent gene activity at the ventral midline presumably reflects lower cell density and/or targeted *Ubx* down-regulation related to midline specification or morphogenesis. The separate posterior terminal spot may relate to proctodeum formation or, given its transience, may be an artefact from the surrounding pattern of down-regulation. As in earlier stages, the anterior boundary of *Tu-Ubx* appears to lie in the Op2 segment, one segment (at least) behind the emerging L4 limb. I carried out double *in situ* hybridisation for *Tu-Ubx* and *Tu-Dll*, which helped confirm both the ectoderm-restricted localisation of *Tu-Ubx*, and the gap between L4 and the anterior *Tu-Ubx* transcription boundary in the opisthosoma (Figure 6.2.3,iii).

ii) *Tu-Ubx* during germband contraction: 45-50hr AEL

As the germband undergoes axial contraction onto the ventral surface, the final larval spider mite body plan is refined, involving extreme reduction of the opisthosoma as well as dorsal closure, gnathal and proximo-distal limb organogenesis (Anderson 1973; Anderson 1999). I detected *Tu-Ubx* at this late stage in a smaller - less broad than before and shorter - ectodermal domain, corresponding to the contracted morphology of the opisthosoma (Figure 6.2.3,iv-v). I observe that the anterior boundary of *Tu-Ubx* transcription seems to have moved forward to the posterior margin of the L4 segment, but given lateral and axial opisthosomal contraction it is not clear if this is a real shift forward within the Op2 or into the Op1 segment (homologous perhaps to the late anterior shift in the *Lithobius Ubx* domain). The posterior *Tu-Ubx* margin is located just anterior to the proctodeum (Figure 6.2.3,v), suggesting that even if the terminal midline 'spot' seen during ventral closure represents a derived proctodeum specification role, it would be very transient; *Ubx* is not generally associated with posterior-most opisthosomal or telson structures. Again, *Tu-Ubx* transcripts are less detectable at the ventral midline of the clear opisthosomal domain, attributable to lower cell density or specific *Ubx* repression associated with maintaining ventral midline fate and morphogenesis (Figure 6.2.3, iv-v).

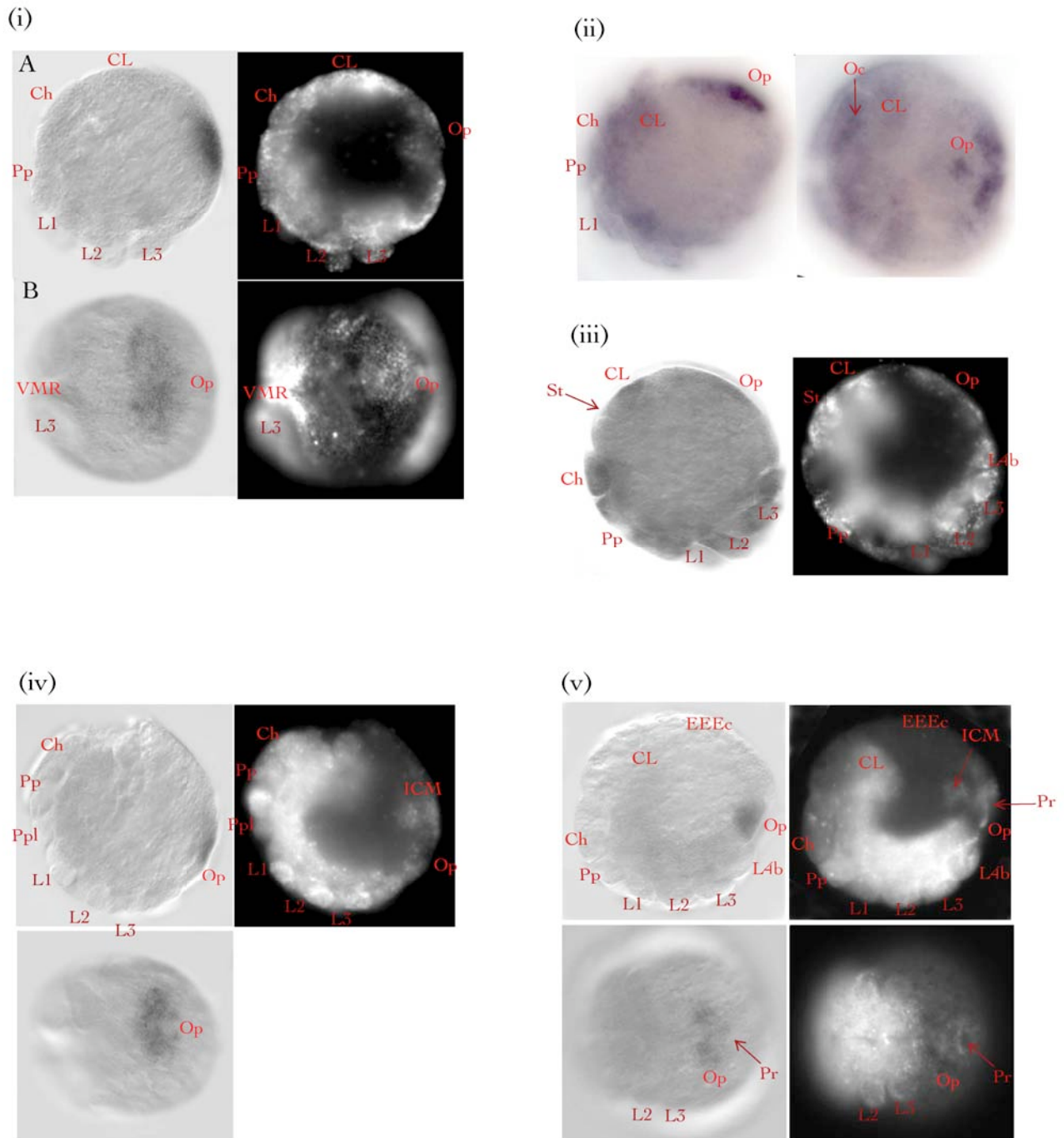


Figure 6.2.3 *Ubx in situ* mRNA expression pattern from ventral ridge closure to germband contraction stage: transcripts persist in restricted opisthosomal domain with straight anterior border in the proximity of the Op2 segment; downregulation at the posterior-most ventral midline is observed at later stages. (i) *Ubx* at ventral ridge closure. (ii) *Ubx* at ventral closure. (iii) *Ubx* + *Dll* double/combined in situ at ventral closure. (iv)-(v) *Ubx* at germband contraction.

CL - cephalic lobe; Op - opisthosoma; ICM - inner (endodermal) cell mass; Ch - chelicera; Pp - pedipalp; Ppl - Pp lobe; L1 - walking limb pair I; St - putative stomodeum invagination region; Pr - putative proctodeum invagination region; Oc - ocular segment/semi-lunar groove; GD - germdisc; EEEc - extra-embryonic ectoderm

6.2.4 Inferences and hypotheses on *Tetranychus* *Ubx/Dll* interactions

The genetic basis of limbless chelicerate and insect tagma?

Previous authors assert that *Ubx* regulation of *Dll* has evolved within the arthropods from a non-repressive interaction in the sister taxon Onychophora to powerful *Ubx* repression of *Dll*, and hence abdominal limbs, in higher insects (Galant and Carroll 2002; Gebelein et al. 2002; Gebelein et al. 2004; Grenier and Carroll 2000; Levine 2002; Pavlopoulos and Averof 2002; Ronshaugen et al. 2002; Vervoort 2002). Exclusive mRNA transcription domains for *Tu-Dll* and *Tu-Ubx* recapitulate the pattern observed in other chelicerates, and indicate a possible direct or indirect repressive interaction between the two genes (Abzhanov et al. 1999; Damen et al. 1998; Popadic and Nagy 2001). However, in the absence of functional data for *Tetranychus* or other chelicerates this inference, being based on circumstantial expression data alone, this idea is restricted to the level of speculation: Could it be that *Ubx* directly represses *Dll* in the chelicerate opisthosoma as well as in the insect abdomen? If so, has the same repression mechanism evolved independently in both sub-phyla in a remarkable instance of parallel evolution, or do distinct genetic mechanisms underlie the convergent phenotypic outcome of posterior limblessness?

Conservation vs. convergence in arthropod Ubx-Dll interactions

As already mentioned in 6.1.2, *Ubx* C' terminal protein sequences are proposed to explain differences in *Dll* regulatory properties of **crustacean** (*Artemia*) and **insect** (*Drosophila*) *Ubx* proteins: repressive *Ubx* is inhibited by S/T phosphorylation sites in *Artemia*, whereas in insect *Ubx*, S/T site loss and replacement by poly-A binding domain releases and amplifies powerful *Dll* repression (Deutsch and Mouchel-Vielh 2003; Galant and Carroll 2002; Ronshaugen et al. 2002). Although this is a simplification given that effects of protein modification depend on target *cis*-regulatory element responsiveness and probably also on cell-specific co-factors, the mechanism is nevertheless sufficient to explain why *Ubx* and *Dll* are co-expressed in the limbed, crustacean pleon, but *Ubx* strongly represses *Dll* throughout the insect abdomen (Galant et al. 2002; Gebelein et al. 2002; Gebelein et al. 2004; Hsia and McGinnis 2003). There is no functional data available to shed light on the nature or significance of evolution in **myriapod** *Ubx* protein or *Dll cis*-regulatory sequences, although the presence of numerous GSK3 or MAPK phosphorylation sites at the C' end of the *Ubx* protein could indicate a mechanism of inhibited repression as in *Artemia* (see Figure 6.1.2d). Throughout the myriapod trunk, *Dll* is co-expressed with *Ubx*, indicating lack of repression between *Ubx* and *Dll* in the myriapod lineage (plus forward movement of the anterior *Ubx* boundary relative to the extreme posterior domain of onychophoran *Ubx*) (Abzhanov et al. 1999; Hughes and Kaufman 2001; Hughes and Kaufman 2002b). Myriapoda branch basal to Pancrustacea and are either (i) sister to, or (ii)

immediately derived from, the basal Chelicerata (Blaxter 2001; Cook et al. 2001; Hughes and Kaufman 2001; Hwang et al. 2001; Kutsche and Burmester 2001; Negrisola et al. 2004). As neither myriapod nor crustacean Ubx can repress *Dll*, and insects likely derive from within Crustacea, the exclusive pattern of *Tetranychus* opisthosomal Ubx and prosomal *Dll* transcription points to a repressive regulatory interaction that would be unique to chelicerates, evolved independently of the insects and operating by a causal mechanism as yet uncharacterised.

A model based on current knowledge

The phylogenetic distribution of i) arthropod Ubx and *Dll* protein/mRNA expression data plus ii) certain hypothetically functional Ubx sequence motifs²⁰ as reviewed above, seems most parsimoniously to support a model based on independent evolution of Ubx-*Dll* repression in Chelicerata and Insecta. Given this hypothesis it would be surprising to find that the exact same genetic mechanism had evolved separately twice, but without further data more solid conclusions cannot be made regarding convergent vs. divergent genetic networks that may underlie phenotypic convergence on limbless posterior tagmata.

6.3 Testing for Ubx/abdA protein activity in *Tetranychus* with antibody FP6.87 (α -UbdA)

6.3.1 The UbdA/FP6.87 antibody

Kelsh et al. (1994) produced a monoclonal antibody (MAb), raised against a carboxy terminal homeodomain + UbdA motif shared by both Ubx and Abd-A proteins in *Drosophila* (de Rosa et al. 1999; Kelsh et al. 1994). The antibody, named FP6.87 after the clone it was purified from, targets the conserved epitope sequence LKKEIQAIKELNEQEK: boxed residues are those essential for antigen recognition and antibody binding (Averof and Patel 1997; Kelsh et al. 1994). Since initial positive results were gained for *Drosophila* and *Schistocerca* Ubx and AbdA antigens, the FP6.87/UbdA antibody has been used to detect combined Ubx/AbdA protein domains in thoracopod segments of a wide range of crustaceans and abdominal segments of numerous apterygote and pterygote, hemi- and

²⁰ Salient coding sequence and Ubx gene expression data for a range of arthropod taxa are summarised in Chapter I section 1.2.2 and graphically represented in Figure 1.2.2a. Section 6.1.2 of this chapter introduces *Tetranychus* Ubx data to the argument, briefly reviewed here in section 6.2.4 with reference to chelicerate vs. insect Ubx-*Dll* interactions.

holo-metabolous insects (Abzhanov and Kaufman 2000a; Averof and Akam 1995; Averof and Patel 1997; Blin et al. 2003; Gullan and Cranston 2000; Palopoli and Patel 1998; Suzuki and Palopoli 2001; Walldorf et al. 2000; Zheng et al. 1999). Even beyond the Pancrustacea, FP6.87 marked nuclear UbdA antigens in trunk segments of *Ethmostigmus* and *Oxidus* (myriapods), terminal lobopod-bearing segments of *Akanthokara* (onychophora), and within Chelicerata, opisthosomal segments of *Limulus* (Xiphosuran), *Paruroctonus* (scorpion), *Cupiennius* and *Achaeearanea* (spiders) (Abzhanov and Kaufman 2000a; Abzhanov et al. 1999; Palopoli and Patel 1998; Popadic and Nagy 2001). The broad diversity of arthropod species positively cross-reactive with MAb FP6.87 directly reflects high sequence conservation within Ubx and AbdA antigenic epitope regions: see Figure 6.3.1 for relevant Ubx sequence comparison across arthropod sub-phyla (de Rosa et al. 1999). As an aliquot of MAb FP6.87 was available to me, I tested it for cross-reactivity in *Tetranychus* embryos, predicting that it would reveal Ubx protein expression – and Ubx alone, as *abd-A* is most likely absent from the *Tetranychus urticae* genome (c.f. section 6.1.4).

6.3.2 Lack of anti-UbdA cross-reactivity in *Tetranychus*

I carried out immuno-staining experiments on spider mite embryos fixed from blastoderm to late limb development stages. Positive control embryos all successfully marked mitotic centromeres *via* an antibody against phospho-Histone-III (Figure 6.3.2,i-ii). However, I did not detect UbdA protein at any stage in embryogenesis (Figure 6.3.2,iii-vii), in spite of testing multiple FP6.87 primary concentrations (1:3, 1:10, 1:50) and developing DIG-AP reactions at room temperature for 5' – 45', or more gradually at 4°C for 30' – 8hrs.

6.3.3 Assessment of negative results based on epitope conservation

Consistent failure to detect Ubx protein expression indicates that:

- i) the epitope targeted by the FP6.87 antibody is divergent in *Tetranychus* Ubx and so cannot effectively cross-react with it, or
- ii) the target epitope is present in *Tetranychus* Ubx but disrupted by embryo fixation conditions, disabling potential detection by FP6.87.

Ubx epitope divergence?

Within the anti-UbdA epitope region of *Tetranychus* Ubx are two divergent amino acid residues: leucine/L → glutamine/Q, and glutamic acid/E → alanine/A (c.f. Figure 6.3.1). These amino acid sites

	1	11	21	31	
Dm-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	INSECTA
Dv-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Tc-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Jc-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Am-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Fc-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Af-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQDKRIT	CRUSTACEA
Sc-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQENNHR	
Ps-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Tc2-Ubx	LCLTERQIKI	WFQNRMMKIK	KETQAIKELN	EQEKQAAQ	
Er-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	MYRIAPODA
Ag-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQDKQAAQ	
Gm-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
La-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Sm-Ubx	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQEKQAAQ	
Cs-Ubx1	LCLTERQIKI	WFQNRMMKIK	KEIQAIKELN	EQERQAAQ	CHELICERATA
Cs-Ubx2	LCLTERQIKI	WFQNRMMKIK	KEAQAIKELN	EQERQAAQ	
Tu-Ubx	LCLSERQIKI	WFQNRMMKIK	KEIQAIKELN	AQSKSSQ	
Ak-Ubx	LCLTERQIKI	WFQNRMMKIK	KEMQTIKDLN	EQEKKQR	
	<-----Helix-3----->		<----UbdA---->		

Figure 6.3.1 Ubx C-terminal protein sequence alignment, showing FP6.87 UbdA antibody epitope binding sites. **RED FONT** – C' homeodomain (inc. α -Helix 3); **PURPLE** - UbdA signature motif; **BLUE** – conserved epitope residue; **YELLOW** – divergent residue within epitope; **bold** - *Tetranychus* Ubx.

Abbreviations: Dm = *Drosophila melanogaster*; Dv = *Drosophila virilis*; Tc = *Tribolium castaneum*; Jc = *Junonia coenia*; Am = *Apis mellifera*; Fc = *Folsomia candida*; Af = *Artemia franciscana*; Sc = *Sacculina carcini*; Ps = *Porcellio scaber*; Tc2 = *Trigriopus californicus*; Er = *Ethmostigmus rubripes*; Ag = *Archispirostreptus gigas*; Gm = *Glomeris marginata*; La = *Lithobius atkinsoni*; Sm = *Strigamia maritima*; Cs = *Cupiennius salei*; Tu = *Tetranychus urticae*; Ak = *Akantzokara kaputensis*.

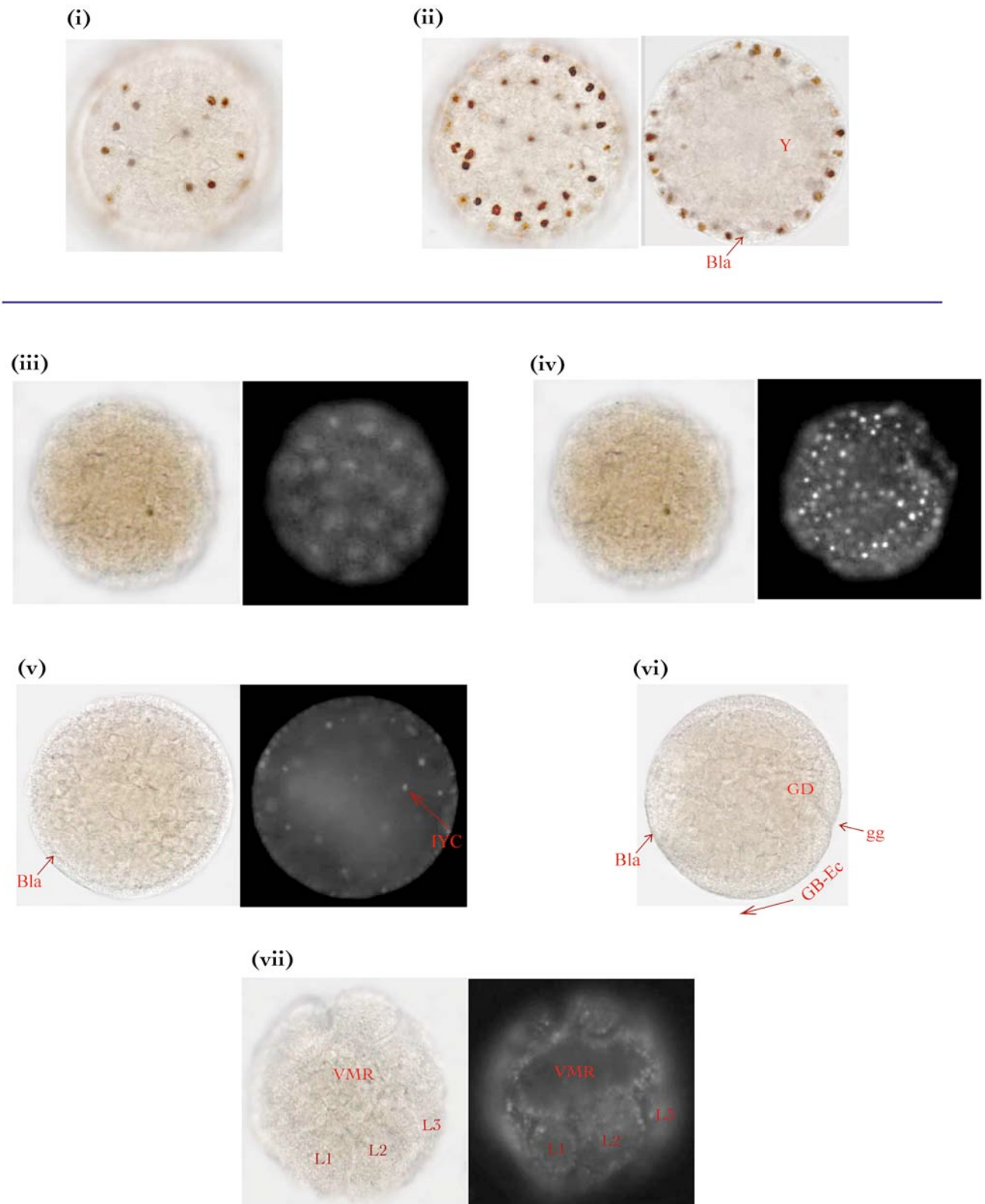


Figure 6.3.2 Antibody staining against phospho-histoneIII (control) and UbdA (anti-FP6.87). (i)-(ii) Control embryos, dividing cells stained positively for pHisIII. (iii)-(vii) UbdA protein not detected at any stage when testing for anti-FP6.87 cross-reactivity: (iii) 64/128-cell stage, (iv) blastoderm, (v)-(vi) germdisc-germband stages, (vii) early limb bud stage.

Bla - blastomere cell (layer); Y - yolk; GD - germ disc; IYC - inner yolk cell; L1 - walking limb pair 1; VMR - ventral midline ridge; GB-Ec - germband ectoderm.

are critical in antibody-epitope binding and hence, especially given the chemical disparity between the substituted and substituting amino acids (L - non-polar, Q - polar, E - acidic, A - non-polar), may well explain failure of MAb FP6.87 to bind within whole mount spider mite embryos (Alberts et al. 1994; Kelsh et al. 1994). A third divergent residue is present within the epitope region as a glutamic acid/E → serine/S substitution, and although this serine is not known to occupy a critical antibody-recognition site its stable, polar –OH functional group may significantly disrupt biochemical properties dependent on the glutamic acid –O⁻ ionised group (Alberts et al. 1994). All other arthropod Ubx epitope sequences are entirely conserved, but onychophoran Ubx is divergent in amino acid sequence at one site within the epitope where alanine (non-polar) is replaced by threonine (polar) – c.f. Figure 6.3.1 (Alberts et al. 1994; Grenier et al. 1997). As UbdA protein has been reported in the posterior lobopod segment of the homonomous *Akanthokara* trunk by detecting for MAb FP6.87 cross-reactivity, the single A→T amino acid change appears not to significantly affect FP6.87 function in recognising Ubx (Grenier and Carroll 2000; Grenier et al. 1997). However, antibody binding to *Ak*-Ubx may be somewhat disrupted by the substitution, so that reported UbdA protein expression in *Akanthokara* may correspond mostly, or wholly, to *Ak*-AbdA; this hypothesis could explain the remarkably posterior boundary of the observed onychophoran ‘Ubx’ domain.

Ubx epitope denaturation?

The epitope for the UbdA antibody in *Tetranychus* Ubx may have been rendered dysfunctional by physical and chemical conditions during fixation (see Materials & Methods, section 8.1.4). Although p-Histone-III and p-Smad-1 antibodies are cross-reactive, indicating that embryonic proteins are not severely disrupted by fixing, the UbdA antibody could have greater sensitivity to the process as each antibody owns specific bio-chemical properties and vulnerabilities.

Conclusions to Chapter VI

i) *Tetranychus Ultrabithorax*

In section 6.1 of this chapter I have presented cloning of ~900bp of *Tu-Ubx*, including the whole coding sequence. Bayesian inference and the presence of conserved Ubx-specific homeodomain and C' terminal Ubx motif residues confirmed the true orthology of this *Hox* gene to *Ubx*. As documented in section 6.2, *Tu-Ubx* ssRNA probes ~500bp to 800bp in length allowed detection of mRNA transcription during embryogenesis, revealing *Ubx* expression in an opisthosomal domain tapering and later contracting laterally and axially in line with extreme reduction of this part of the spider mite body. I have suggested that - as far as can be ascertained - the straight anterior *Tu-Ubx* mRNA boundary correlates with the second opisthosomal segment (Op2), agreeing with previous proposals based on horseshoe crab, scorpion and spider Ubx, that the ancestral chelicerate anterior Ubx expression boundary is in Op2. *Tu-Ubx* is clearly not expressed in most posterior body regions, the final posterior *Tu-Ubx* expression boundary lying before the proctodeum and telson. Unfortunately I could not augment *Tu-Ubx* mRNA transcript detection with an account of *Tu-Ubx* protein dynamics, as Kelsh's FP6.87/ α -UbdA antibody did not cross-react in *Tetranychus urticae*, in spite of its broad cross-reactivity in other arthropod species. I have proposed that lack of cross-reactivity in spider mites most likely stems from epitope divergence, evident when amino acids of the Ubx+AbdA C' epitope are compared: unusually, *Tu-Ubx* contains three substitutions within this generally highly conserved region, two of which are known to critically affect antibody-epitope recognition.

In the context of comparative *Ubx* coding sequence and gene expression data for a range of taxa from each major arthropod sub-phylum, I have considered how *Tu-Ubx* data may contribute to resolving the issue of whether exclusive opisthosomal BX-C and prosomal *Dll* domains in chelicerates reflect any convergence or conservation in gene function relative to BX-C-mediated *Dll* repression in the abdomen of higher insects such as *Drosophila*. Examination of Ubx C' terminal motifs and *BX-C* vs. *Dll* expression in many arthropods (including *Tetranychus*) most parsimoniously points to **independent** evolution of BX-C factor-mediated *Dll* regulation in chelicerate and insect lineages, with disparate genetic mechanisms responsible for repressed *Dll* transcription and phenotypic convergence on limbless posterior opisthosomal or abdominal segments. Lack of functional data seriously hinders further hypothesis testing at this point, as without it we cannot know if *Tu-Ubx* directly represses *Dll* in the opisthosoma or if the two genes are coincidentally expressed in strictly exclusive domains, suggestive of repression but not necessarily demonstrative of it. Nevertheless, with respect to previous inferences about the evolution of Ubx-*Dll* repression based on Ubx C' terminal amino acid

changes, broader taxonomic comparison – significantly including *Tu-Ubx* – reveals greater protein diversity than anticipated, complicating and casting doubt on evolutionary models of Hox function based on fewer taxa. I have argued that multiple mechanisms for *Dll* repression by BX-C genes may exist, potentially mediated by transcription factor changes, but ultimately critically dependent on *cis*-regulatory responsiveness of target genes to permit functional effects and affect convergent or divergent phenotypic outputs.

ii) *Tetranychus abdominal-A*

Section 6.1.4 has outlined difficulties encountered in amplifying a *Tetranychus abdA* ortholog from genomic DNA and cDNA templates, in spite of testing many PCR programs and conditions as well as combinations of degenerate primers targeted against conserved homeobox and N' terminal hexapeptide motif coding sequences. It appears that *Tu-abdA* is missing from an otherwise typical and complete *Tetranychus Hox* cluster, and so further consideration of *abdA* as a potential limb suppressor is irrelevant for this spider mite species. I have proposed that the absence of a *Tu-abdA* ortholog from the spider mite genome may be an example of gene loss associated with secondary reduction of the body plan, namely an extreme opisthosomal contraction. I have argued that in support of this hypothesis, parallel losses of *abdA* from the oribatid mite *Archegozetes longesitosus* genome, and clearly independently from the crustacean *Sacculina carcini*, can also be linked with secondary body plan reduction, respectively affecting miniaturisation of the mite opisthosoma and vestigial crustacean abdomen.

CHAPTER VII

DISCUSSION



Mygale de Sauvage

7.1 Assessment of conservation within the *Dll* genetic regulatory network for early limb development

The *Distal-less* genetic regulatory network (GRN) operating in *Drosophila melanogaster* to specify limb primordia appears to have only very limited points of direct conservation with respect to other arthropods. However, among arthropods and even in more disparate groups (e.g. vertebrates), there are interesting parallels in certain patterning-specific roles carried out by genes within the GRN that may hint at deep homology in genetic elements being used in limb development. With reference to the Aims 3 to 6 listed²¹ in Chapter I, the following paragraphs firstly summarise findings arising from *Tetranychus* as they relate to our picture of conservation and divergence in genetic networks effecting arthropod limb specification, and secondly detail a few hypotheses relating to deep genetic homology.

7.1.1 Antero-posterior specification of limb primordia

Transcription of *Drosophila wingless* (*wg*) is activated by En-mediated signalling from posterior parasegment stripes, such that the secreted signalling molecule Wg is first expressed in stripes anterior to the parasegment boundary. This parasegmental Wg signal is strictly required for embryonic limb primordium specification (Gilbert 2000). In other insects (e.g. *Tribolium*, *Oncopeltus*, *Gryllus*) *wg* loss-of-function has revealed variable requirements for *wg* in appendage specification, but it is as yet unknown whether *Drosophila wg* exemplifies the ancestral state for hexapods or not (Angelini and Kaufman 2005a; Angelini and Kaufman 2005b; Miyawaki et al. 2004). As presented in Chapter IV, I was unfortunately unable to clone and express *Tetranychus* homologs of *wg/Wnt1* or *engrailed*. However, posterior parasegmental expression of one *Tu-en* paralog has been reported by other workers, providing preliminary evidence in support of conserved molecular interactions at parasegment boundaries of the spider mite embryo. Parasegmental *wg* and *en* expression is reported throughout the arthropods, suggestive of a possible conserved molecular mechanism for activation of *Dll* in limb primordial cells along the antero-posterior axis (see Figure 7.1.1) (Damen 2002; Prpic 2004a; Prpic 2004b). Functional studies in *Tetranychus urticae* as well as other non-insect arthropod

²¹ **AIM 3** To clone the *Tetranychus Dll* homolog and to verify its conserved expression during limb development; transcription is predicted from limb specification onwards. **AIM 4** To clone *Tetranychus* orthologs for candidate genes potentially involved in specification of limb primordia *via* direct or indirect regulation of *Dll* – i.e. *wingless*, *engrailed*, *dpp*, *EGFR*, *Ubx*, *abdA* and *Sp*. **AIM 5** To determine mRNA and/or protein expression patterns for these candidate *Dll*-regulatory genes during embryogenesis, with particular attention to early limb development. **AIM 6** To compare sequence data and expression profiles of the *Tetranychus* candidate genes with data available for orthologs of these genes in other arthropods and in more distantly related species where relevant, thereby to assess the possible nature of conservation or divergence in gene regulatory networks operating during arthropod limb specification, consistently marked by *Distal-less* expression.

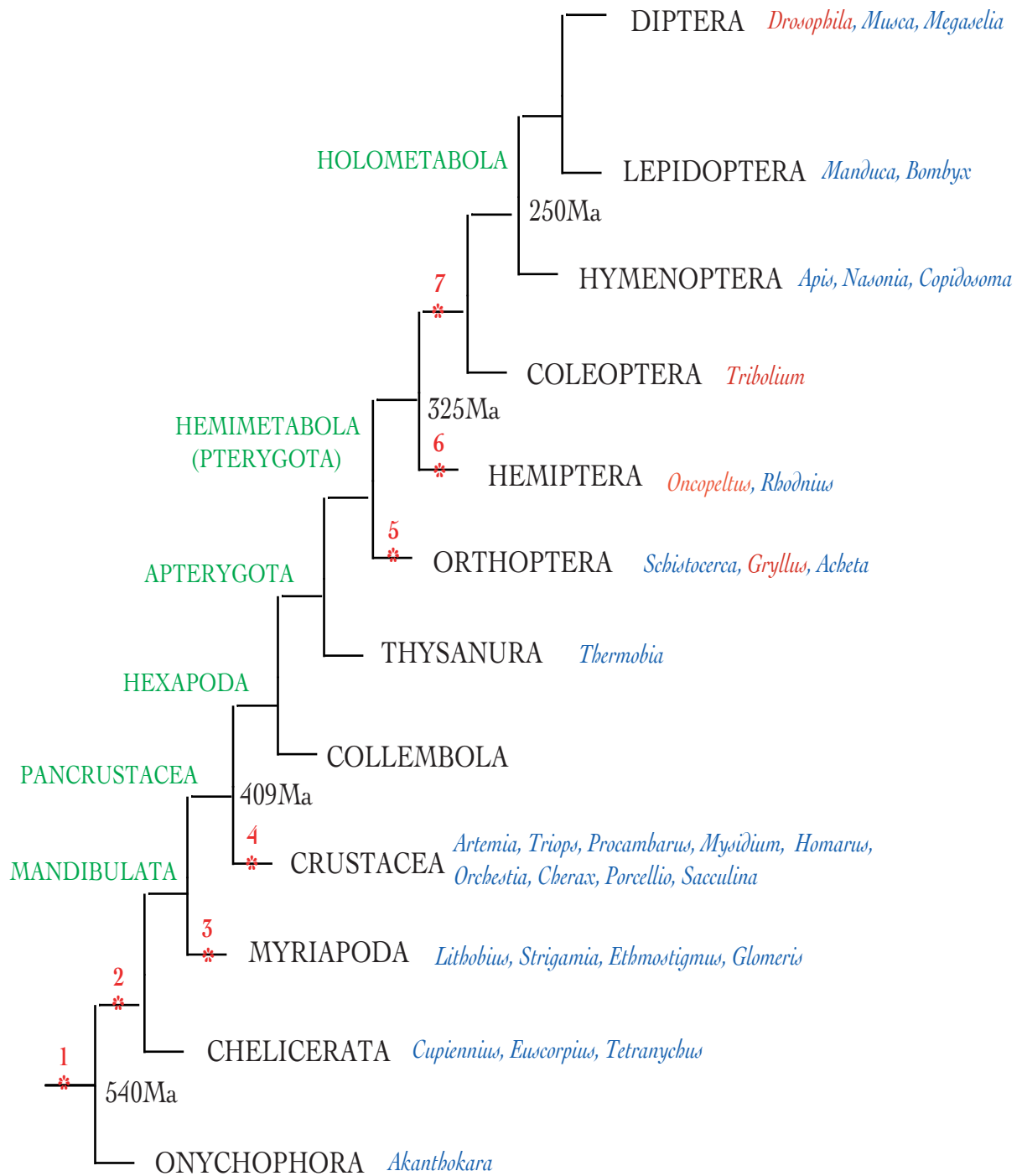


Figure 7.1.1 Possible evolutionary scenarios to explain variable *en/wg* expression and function in metameric patterning and limb allocation in arthropods.

1: Engrailed known to be expressed in conserved metameric pattern; role in limb specification unknown.

2: Conserved *engrailed* and *wingless* gene activity at parasegment boundaries; function in limb allocation not proved.

3: Lineage-specific dissociation of typical ventral para-segmental En/Wg activity from dorsal En/Wg e.g. *Glomeris marginata* (diplopod)

4: Lineage-specific loss of para-segmental Wg/Wnt1 gene expression e.g. *Mysidium colmbiae*

5 + 6: Species-specific evidence that *wingless* is not required for limb allocation - ancestral state for hexapods?

7: Wingless specifically required for limb allocation and patterning e.g. *Drosophila*

Blue - species for which En/Wg data are available. Red - species with available En/Hh/Wg functional data.

taxa are still required to confirm or refute this possibility; the variability in insect requirements for *wg* certainly emphasise this experimental need. Spatio-temporal variation in *wg* and *en* expression also points to the relevance of characterising gene function separately for both of these putative *Dll* GRN players; insect data imply that regulatory conservation may have been more strict with respect to *en* than *wg* interactions, and such variability is worth testing outside the insects in the search for ancestral genetic pathways of *Dll* activation.

7.1.2 Dorso-ventral specification of limb primordia

In *Drosophila melanogaster*, the dorsal and ventral extent of primordial limb fields are delimited by repression of the early *Dll-504* enhancer element in response to graded dorsal Dpp-mediated signalling and a gradient of EGFR-mediated signalling from the ventral midline. Evidence presented in Chapter V suggests that neither Dpp nor EGFR-mediated signalling mechanisms are conserved in *Tetranychus urticae* prosomal limb specification, and comparison with other arthropods confirms the hypothesis that dorso-ventral restriction of *Dll* transcription by Dpp and EGFR gradients is an atypical genetic regulatory mechanism, not conserved outside *Drosophila* (see Figure 7.1.2) (Goto and Hayashi 1997; Kubota et al. 2000; Panganiban 2000).

Decapentaplegic

In the beetle *Tribolium castaneum*, *Tc-dpp* RNA interference does not affect appendage formation, indicating that deployment of *dpp* to define limb primordial *Dll* territories is a mechanism specific to the dipteran *Drosophila* but not to less derived, or ancestral, insect groups (Jockusch and Ober 2004). This conclusion is further supported by observations of Dpp-mediated signalling in spider embryos, although spider *dpp* RNAi unfortunately does not produce a viable phenotype with respect to analysis of possible limb specification roles, because loss of *dpp* function disrupts embryonic axis formation at even earlier stages (Akiyama-Oda and Oda 2006). In the spiders *Achaearanea tepidariorum*, *Cupiennius salei* and now also in the spider mite *Tetranychus urticae* (Chapter V), *dpp* signal transduction is associated with faint, broad prosomal segment domains, followed by persistent activity throughout limb primordia, restricted primarily to the distal tip - with a uniform dorso-ventral distribution (Akiyama-Oda and Oda 2006; Prpic and Damen 2004). A similar distal tip domain is reported for *dpp* ortholog gene expression and signal transduction in the myriapod *Glomeris marginata* and a number of insects (e.g. *Gryllus bimaculatus*, *Schistocerca americana*, *Tribolium castaneum*), but no earlier potential *Dll*-restricting dorsal Dpp signalling domains are reported (Giorgianni and Patel 2004; Prpic 2004a). Considering all comparative *dpp* data available for chelicerates and other arthropods, it appears that the Dpp gradient operating to specify the dorsal limit of *Drosophila* embryonic limb primordia by mediating *Dll* repression is a mechanism unique to *Drosophila* alone.

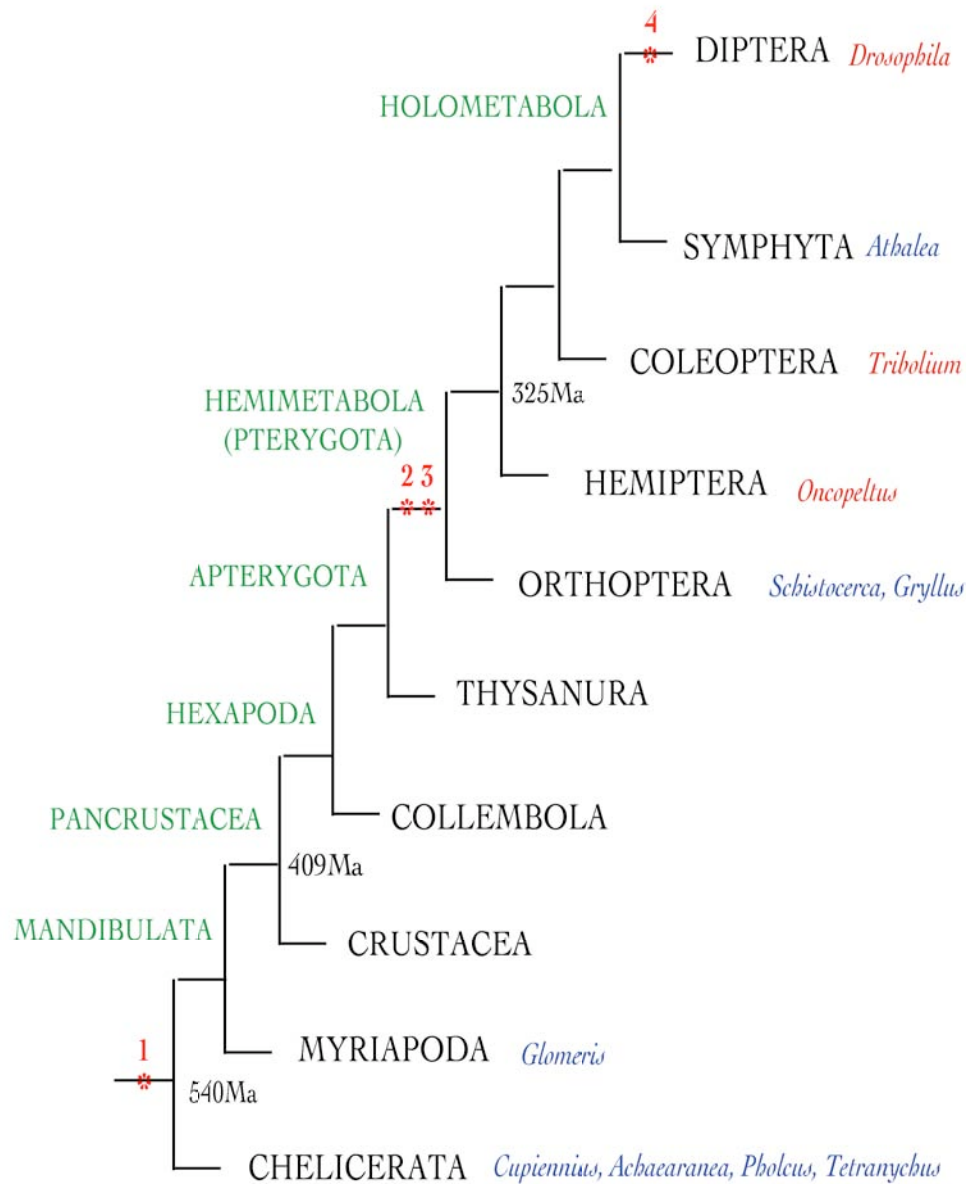


Figure 7.1.2 Possible evolutionary scenarios to explain variable *dpp* expression and function in arthropod limb specification, proximo-distal growth/patterning and dorsal fate specification.

- 1: *Dpp* signaling associated with distal limbs - possible ancestral role in proximo-distal growth/patterning. Low level *Dpp* signal in early segmental stripes, possibly a general activator (of limb specification)?
- 2: *dpp* expression associated with posterior germband structures; unknown function.
- 3: *dpp* expression and function required during dorsal embryo/extra-embryonic fate specification.
- 4: Lineage-specific deployment of dorsal *Dpp* gradient to repress and restrict *Dll* during limb primordium positioning, e.g. *Drosophila melanogaster*.

Blue - species for which En/Wg data are available. Red - species with available En/Hh/Wg functional data.

Traits of Dpp signalling shared in early arthropod limb development indicate a possible ancestral role in inducing outgrowth of the primary limb primordia, and a more clear ancestral role in specifying distal limb fields and orchestrating proximo-distal limb patterning.

Epidermal Growth Factor Receptor

Tetranychus urticae *EGF-R* (*TER*) transcription (Chapter V) reveals gene activity correlated with ventral neurogenesis and patterning, anterior CNS specification, morphogenesis of the brain and visual system, and possible spider-mite specific roles in late proximo-distal limb patterning (Shilo 2003; Shilo 2005). During phases of limb specification and outgrowth, *TER* transcripts are detected in a stripe restricted to the ventral neurogenic midline, with no *TER* activity associated with or reaching to more ventro-lateral regions where limb primordia develop. I conclude that a *TER*-mediated gradient of EGF signalling is not operating during spider mite limb specification, and that ventral limits of *Tu-Dll* transcription must be prescribed through the activity of other, as yet uncharacterised genes. Hence, the EGFR-mediated mechanism operating in *Drosophila* to specify appendage primordia is not conserved in chelicerates, although lack of data for any other arthropod EGFR orthologs means that wider taxonomic sampling is needed before attempting to assert ancestral genetic regulatory states. On the basis of commonalities between the derived species *Drosophila* and *Tetranychus* it seems safe to conclude a conserved and likely ancestral role for EGFR in arthropod ventral and anterior neurogenesis, but regarding limb development little can be said at this time (Mayer and Nusslein-Volhard 1988; Shilo 2005).

7.1.3 Deep genetic homology in limb development?

Animal appendages have evolved independently in the major phyla (e.g. Arthropoda vs. Chordata), resulting in analogous structures whose superficial morphological similarities reflect a common pressure on disparate body plans to meet the demand for locomotion and feeding. Although this means that particular anatomical features of disparate animal appendages cannot be homologised, remarkable apparent conservation of molecular genetic components driving certain events in limb development in a range of phyla, has fed the notion that there may yet be underlying ‘genetic homology’ (Wilkins 2002). It is possible, however, that molecular strategies shared, for example between insects and vertebrates, represent chance convergence due to limitations within ancestral regulatory gene networks leading also to a finite number of terminal differentiation gene batteries and phenotypic outcomes (Davidson 2001; Martinez Arias and Stewart 2002). Selected examples of genetic limb developmental strategies shared between a number of Bilaterian taxa are given below, lending compelling credence to the first hypothesis, that a degree of true ‘deep’ genetic homology may exist:

- *Dll/Dlx* genes are expressed in ectodermal outgrowths of all animal phyla thus far investigated, indicative of inducing some form of outgrowth in an Urbilaterian ancestor (Panganiban et al. 1997; Panganiban and Rubenstein 2002). Frequent correspondence of *Dll* with sensory structures, and identification of distinct regulatory elements for sensory and appendage-related *Dll* transcription, has led to the conclusion that its original ancestral function was in neural patterning, limb functions probably acquired secondarily by regulatory co-option (Panganiban and Rubenstein 2002; Wilkins 2002). Dissociation of neural and appendage inducing regulatory inputs could provide flexibility in genetic regulatory resources such that *Dll* expression could be modulated in specific patterning contexts: Given the unusual anterior sensory appendages found on certain fossil arthropods (e.g. *Kerygmachella*), perhaps *Dll* was recruited from an anterior sensory-appendage associated role to become deployed segmentally via a reiterated neural regulatory sub-routine (Akam 1998; Budd 2002).
- *Sp8* gene homologs in insects (e.g. *Drosophila*, *Tribolium*) and vertebrates (e.g. mouse) are all involved in inducing and maintaining signalling during proximo-distal limb outgrowth,²² suggesting possible deployment in Urbilaterian body wall outgrowths (Beermann et al. 2004; Estella et al. 2003; Triechel et al. 2003).
- Hedgehog signals are active at the antero-posterior boundary of both insect and vertebrate limb primordia, in the form of posterior parasegmental Hh in arthropods (e.g. *Drosophila*, *Artemia*, *Euscorpius*) and vertebrate Shh morphogen synthesis in the posterior compartment of the apical ZPA (zone of polarising activity) (Simonnet et al. 2004; Wilkins 2002). This suggests deployment of Hh morphogens during antero-posterior domain specification in some ancestral limb structure, or co-option of an ancestral antero-posterior patterning program to new sites of limb outgrowth.
- With respect to Wnt pathway activation during limb development²³, vertebrate Wnt signalling is associated with dorsalisation, and *Drosophila* Wnt signalling with ventralisation, of the limb primordium prior to proximo-distal outgrowth: considering the hypothesis of dorso-ventral axis inversion between protostomes (e.g. arthropods) and deuterostomes (e.g. vertebrates), this could betray conserved *Wnt* deployment in assigning dorso-ventral identity, but *after*

²² After initial limb specification, mouse *Sp8* (*mBtd*) acts downstream of Wnt and FGF signals to maintain En and BMP signals, D/V apical ectodermal ridge (AER) patterning and proximo-distal outgrowth (Bell et al. 2003; Kawakami et al. 2004; Triechel et al. 2003; Beermann et al. 2004). The *Drosophila* *Sp8* gene, *buttonhead* (*bt*), acts redundantly with its paralog '*Sp1*' and is expressed throughout ventral thoracic limb primordia, marking proximal to distal limb domains of imaginal discs. *Dm-btd* seems to act downstream of *wg*, *dp* and *BX-C* gene inputs during ventral limb specification, activating *Dll* in parallel; when expressed ectopically in dorsal limb domains, *Dm-btd* can activate *en*, *wg* and *dp* to induce a full genetic program for proximo-distal patterning of a ventral-type limb (Estella et al. 2003).

²³ Intercellular Wnt pathway activation in early embryo axis specification may have been conserved since Cnidarian-Bilaterian divergence (e.g. *Hydra*, *Nematostella*, *Lytechinus*, *Ciona*) (Hobmayer et al. 2000; Imai, 2003; Weitzel et al. 2004; Wikramanayake et al. 2003).

initial limb specification (De Robertis and Sasai 1996; Martinez Arias and Stewart 2002; Wilkins 2002; Yabe et al. 2003; Zeng et al. 2000).

- Antagonism between dpp and Sog affects ventral specification (and ventral CNS induction) in arthropods (e.g. *Drosophila*, *Achaearanea tepidariorum*), and similar antagonism between the vertebrate homologs BMP2/4 and Chordin is responsible for dorsalisation and dorsal nerve cord specification. This indicates a deeply conserved genetic interaction affecting the secondary body axis, manifest in arthropods and vertebrates on opposite sides of the dorso-ventral axis due to inversion (Akiyama-Oda and Oda 2006; De Robertis and Sasai 1996; Finnerty et al. 2004; Holland 2004; Wilkins 2002).
- TGF- β mediated signalling by Dpp in arthropods generally activates distal limb outgrowth, whereas BMP in vertebrates negatively regulates distal outgrowth during digit vs. interdigit morphogenesis. An opposing TGF- β regulatory role in an analogous (i.e. distal limb) domain may be linked to underlying genetic homology, but this – and the following examples of trends in RTK signalling – is speculative, and implies extensive divergence in allied regulatory circuitry within each lineage (Martinez Arias and Stewart 2002; Wilkins 2002).
- RTK signalling *via* EGF in *Drosophila* and FGF in vertebrates is graded during distal limb development in both, possibly indicative of a conserved proximo-distal genetic routine (Campbell 2002; Galindo et al. 2002; Tickle 2003). Use of an RTK signalling gradient for morphogenesis may alternatively represent convergence, especially given different RTK activation by EGF vs. FGF ligands, and the lack of apparent EGFR activity gradients associated with distal *Tetranynchus* limb morphogenesis (c.f. Chapter V).
- Vertebrate EGFR signalling *via* ectodermal EGF and TGF- α appears to be responsible for AER induction, with EGF specifying ventral identity at the limb apex and antagonising *Bmp4* transcription during proximo-distal development (Dealy et al. 1998). *Drosophila* Dpp (*Bmp2/4*) signalling is also antagonised by EGFR in the limb, but proximo-distal patterning relies not on ventral EGFR but a distally localised gradient, indicating limited genetic conservation (Bridge et al. 2000; Rutledge et al. 1992). Lack of ‘deep’ ventral limb conservation holds true for comparisons between basally diverging arthropod lineages, as spider mite EGFR activity is restricted to ventral mid-line and anterior neurogenic regions rather than being linked to limb specification or early proximo-distal growth. The possibility of a general, conserved genetic sub-routine for EGF-mediated Dpp/BMP antagonism cannot be ruled out, however, having been characterised in both *Drosophila* and vertebrate genetics.

7.2 The limits of comparative gene expression studies

Conclusions based on gene expression studies alone must always carry the caveat that they are subject to revision in the light of demonstrations of actual gene function, by loss-of-function (e.g. by RNAi) or ectopic genetic mis-expression. This limitation has been encountered at several junctures in this thesis, first in the form of uncertainty in hypotheses that could be sound but require functional verification, and secondly where functional studies in other species have thrown asunder previous notions connecting particular gene expression patterns to strong assumptions about conserved gene function. A few examples arising from investigations of putative *Dll* regulatory genes in *Tetranychus* and other significant arthropod species, are detailed below:

7.2.1 Example A: Affinity of the arthropod labrum

Debates about the labrum centre on whether it is segmentally derived and whether it is truly appendicular. The spider mite labrum is a bilobed structure in which *Tu-Dll* is expressed bilaterally, having apparently resolved from earlier gene expression domains in the ocular segment. Furthermore, *Tu-Dll* is expressed in presumptive distal territories within labrum lobes, indicating that the structure represents a ‘whole’ limb, or telopod. *Tu-Dll* gene activity therefore supports a segmental appendage affinity for the chelicerate labrum, and by inference the arthropod labrum. Strengthening this conclusion with data from another mite, bilateral clusters of *Dll*-positive cells also appear to move from ocular lobes to the labrum in embryos of the oribatid *Archegozetes longesitosus* (Thomas and Telford 1999). Functional studies in other species have revealed that the labrum is of appendicular affinity, but indicate flexibility in which proximo-distal limb segment derivatives form the labrum structure itself. For example, the homeotic *Tribolium* mutant *Ag⁵* indicates that the beetle labrum represents a fused pair of endites, derived from most proximal – coxopod - limb segments, but the spider labrum is derived from a whole limb, having telopod identity which is abrogated by *Cs-Dll* RNAi (Haas et al. 2001a; Popadic et al. 1998; Schoppmeier and Damen 2001). *Tetranychus Dll* expression tentatively indicates a whole-limb labrum, suggesting that this state is ancestral at least for chelicerates; however, without a *Tu-Dll* loss-of-function phenotype we cannot prove conclusively that *Dll* is acting in an appendage patterning role in the spider mite labrum, nor can solid models be proposed regarding ancestral character states.

7.2.2 Example B: Sub-functionalisation among *engrailed* paralogs

Diverse forms and degrees of functional divergence between *en* paralogs have been documented in several insects, crustaceans and the spider *Cupiennius* based on comparative gene/protein activity domains, indicating a tendency for functional specialisation or sub-functionalisation between paralogs (e.g. in segment polarity and neurogenesis roles). In the insects *Drosophila*, *Bombyx*, *Schistocerca* and *Thermobia* there are subtle temporal differences in appearance of anterior segmental stripes, and in *Drosophila*, *engrailed* and its paralog *invected* are both required proper for A-P wing disc patterning (Dhawan and Gopinathan 2003; Peel et al. 2006; Peterson et al. 1998). In crustaceans *Porcellio*, *Procambarus* and *Sacculina*, both paralogs are expressed in ectodermal stripes and appendage domains early on, followed by separate expression in ventral neurogenic cells (*Dm-en* ortholog) vs. posterior segmental structures (*Dm-inv* ortholog) (Abzhanov, 2000; Quéinnec, 1999; Gibert, 2000). Finally, *Cupiennius salei* *En-1* and *En-2* share similar later expression in parasegmental stripes and posterior appendage domains, as well as at the proctodeum and anterior to the labrum, but establishment of parasegmental expression appears to be regulated very differently in the two paralogs; *Cs-En1* stripes arise in antero-posterior progression, but *Cs-En2* stripes appear later, and generate periodically as ‘doublet’ stripes from a posterior proliferation zone. *Cs-En2* is also observed in the neurectoderm and stomodeum, unlike *Cs-En1* (Damen 2002). As yet none of these cases of apparent spatio-temporal divergence in arthropod *engrailed* paralog regulation have been tested by gene knockouts targeting individual paralogs, so the connection between gene expression and functional divergence is not yet proved. Turning to *Tetranychus*, one of the two spider mite *engrailed* paralogues, *Tu-en1* or *Tu-en2*, definitely exhibits ‘typical’ posterior segmental appendage expression, but in order to meaningfully address possible paralog sub-functionalisation it remains necessary to characterise *Tu-en1* and *Tu-en2* expression patterns separately, and then obtain separate loss-of-function phenotypes to detect any *en* regulatory divergence that impacts upon gene function as well as expression.

7.2.3 Example C: *Dll* repression in the chelicerate opisthosoma

As mentioned in Chapter VI, *Ubx* and *abdA* activity is consistently located in the limbless chelicerate opisthosoma, in segments caudal to a conserved anterior boundary in the second opisthosomal (Op2) segment. This statement is based on detection of mRNA transcripts and/or protein domains in embryos of a wide taxonomic range of chelicerates, including Xiphosura, scorpions, spiders and now the mite *Tetranychus urticae* (Damen et al. 1998; Popadic and Nagy 2001). In direct contrast, *Distal-less* activity during early limb specification is restricted to primordial limb fields within prosomal segments, creating a situation in chelicerates reminiscent of that in higher insects, in which thoracic *Dll* activity is restricted by Bithorax complex protein activity in the abdomen (Panganiban 2000).

Higher insect *Dll* transcription is inhibited throughout the abdomen due to direct repression by Ubx and AbdA, with associated co-factors (Gebelein et al. 2004; Levine 2002). Given the large phylogenetic disparity between insects and chelicerates, and the existence of intermediate lineages in which Ubx, AbdA and Dll show non-repressive interactions, it would be very surprising to find that chelicerate Ubx and AbdA homologs repress *Dll* in the opisthosoma, or if they do, repress *Dll* by the exact same mechanism as in insects. If chelicerate Ubx and/or AbdA do repress *Dll* in the opisthosoma, independent co-option would seem most probable, involving recruitment of posteriorly-expressed *BX-C* genes to mediate limb suppression *via* genetic networks and/or regulatory elements that are divergent relative to insects but happen to converge on the output of silencing *Distal-less* transcription. Without any functional data for chelicerate *Ubx* and *abdA* homologs, we cannot progress beyond mere speculation: to prove the hypothetical scenario of convergence on *BX-C*-mediated *Dll* repression, there is a need to show that without *Ubx* and/or *abdA* activity, *Dll* transcription is permitted in the opisthosoma, and with *Ubx* and *abdA* ectopically expressed in the prosoma, *Dll* is suppressed.

7.2.4 Example D: Variable interactions and dissociation between segment polarity gene expression and function in limb specification

Limited circumstantial evidence points to flexibility in genetic networks activating *Dll* in limb primordia with respect to how segment polarity gene inputs from Wg and En signalling take effect. As mentioned in Chapter IV, in early *Schistocerca gregaria* segmentation, *wingless* stripes appear before *engrailed*, indicating at least a degree of regulatory divergence from the scenario in *Drosophila* where En signalling first activates Wg (Dearden and Akam 2001). Even more compelling is evidence from the crustacean *Orchestia cavimana*, showing Dll protein in segmental limb primordia *before* parasegmental *en* and *wg* gene expression are observed (Hejnol and Scholtz 2004). These cases imply non-*Drosophila*-like regulation of Wg by En at least at some points in development, and imply that early *Dll* *cis*-regulatory elements are targeted by transcription factors other than those induced by En/Wg activity, similarly unlike the situation in *Drosophila*. Recent functional studies have brought even more lucidity to this topic, and offer a degree of explanation in that in some species at least (e.g. *Oncopeltus* and *Gryllus*), parasegmental *en/wg* transcription need not betray a role in antero-posterior limb specification. In the hemipteran *Oncopeltus fasciatus* RNAi phenotypes indicate *Of-en* function in segmentation and appendage specification, but in spite of parallel gene expression, *Of-wg* only functions in early segmentation (Angelini and Kaufman 2005a). In the orthopteran *Gryllus bimaculatus*, *Gb-wg* and *Gb-bb* RNAi have no clear phenotypes, also casting doubt on the function of En and Wg signalling, although the authors infer failed gene knockout or possible *Wnt* redundancy (Miyawaki et

al. 2004). In conclusion, by correlating conserved segment polarity gene expression with functions identical to the *Drosophila* paradigm, the importance of arthropod *en* and *wg* interactions in *Dll* activation appears to have been surprisingly over-estimated.

7.2.5 Verdict on genetic inferences based on *Tetranychus urticae* data

At present there are no practical means to demonstrate candidate gene function in the spider mite (c.f. section 7.3), and within the scope of this project I could not develop and test gain- or loss-of-function assays with respect to genes in the putative GRN for limb specification. Therefore, in addressing evo-devo hypotheses with data from spider mites, we are constrained to making inferences based on mRNA transcript and/or protein activity domains. In the light of a few clear demonstrations of gene function in other chelicerate taxa, such inferences may arguably carry a good degree of weight due to likely conservation by phylogenetic association. However, recent surprises brought from RNAi gene knockouts in relatively closely related insect species are a reminder of the need to exercise caution with comparative inferences, and make final genetic functional analysis the goal and standard for future hypothesis testing in evo-devo. The construction of neat evolutionary scenarios based on gene expression patterns in a broad taxonomic range of species is a very valuable first step, but any tidy models must be just as ready to be confirmed as radically re-organised or thrown into disarray as they accommodate emerging, exciting data on gene function that promises to add more power and complexity to our understanding of the evolution of developmental genetics.

7.3 Assessment of *Tetranychus urticae* as a model chelicerate

In Chapter I (section 1.3.2) I made a preliminary assessment of the potential suitability of *Tetranychus urticae* as a species for molecular evo-devo studies, and in this section I review this assessment according to a) the results of thesis Aims²⁴ such as the development of a new, reliable protocol for

²⁴ **AIM 1** To describe embryonic and post-embryonic development in the spider mite *Tetranychus urticae*, with a particular emphasis on limb formation. **AIM 2** To develop and demonstrate a working protocol for whole mount *in situ* hybridisation and antibody staining in *Tetranychus*. **AIM 7** To assess the validity and practicality of *Tetranychus urticae* as a model species for evolution and development studies.

embryo fixation to permit detection of gene activity in spider mite embryos, combined with b) the recent news regarding progress on genomic and gene functional data collection.

7.3.1 Critique of practical workability

I found that spider mites, being plant pests, thrive well and are cultured easily in very large numbers when kept living on young French bean plants under normal lab environmental conditions. Eggs are easily removed from leaves and collected separate from adults and nymphs by using 200µm primary and 70µm-mesh secondary sieves. Once collected, early cleavage stage zygotes are clearly identifiable under a dissecting microscope by virtue of dense cytoplasmic haloes around nuclei, thus making precisely timed developmental studies very feasible. The detailed overview of *Tetranychus* embryonic and post-embryonic development presented in Chapter II includes time scales of early cleavage divisions and subsequent morphogenesis, proving a framework (c.f. Aim 1) to locate subsequent experimental embryos within well restricted windows of time After Egg Laying – i.e. hours AEL. Therefore, with respect to ease of culture, embryo collection and identification of stages of embryogenesis, *Tetranychus urticae* is an ideal ‘model’.

While developing efficient protocols for *in situ* mRNA hybridisation and antibody staining of whole-mount spider mite embryos (c.f. Aim 2), I had to overcome many difficulties and in the end abandoned the ‘mass method’ documented by Dearden et al. in their papers on germ line specification and early segmentation in *Tetranychus* (Dearden et al. 2003; Dearden et al. 2002). See **Appendix 7** for a detailed criticism of these papers, based on an awareness that the more precise understanding of development now available illuminates certain inaccurate developmental interpretations and methodological deficiencies, both impacting upon their interpretations of germ cell fate and segmental register. The method of Dearden et al. gives an allegedly ‘adequate’ <1% success rate regarding embryos stained positively for specific transcripts or proteins, but when tested – even with numerous basic protocol modifications - I found this fraction to be far, far lower and either way unsuitable when seeking to examine unknown gene expression patterns, as it is impossible to distinguish junk signal and general mess from putatively gene-specific signals. The fixation protocols I went on to develop (c.f. Chapter VIII: Material & Methods) provided a 100% success rate regarding detection of target mRNA or protein in staged embryos, but the procedure itself was very laborious, time-consuming and required great dexterity, so I consider it far from ideal as an option for large-scale or long-term embryo developmental investigations. Furthermore, as discussed throughout results Chapters III – VI in response to mysteriously negative results, *Tetranychus* embryos were sometimes inexplicably un-responsive to mRNA probes and often un-cross-reactive with antibodies tested.

7.3.2 Assessment of progress on genomics and gene function

The small genome size of *Tetranychus urticae* (approx. 75Mbp) makes it a suitable species for full genome sequencing. The US Department of Energy Joint Genome Institute (DOE-JGI) have agreed to sequence the spider mite genome, and the current status of this enterprise is that DNA is waiting to enter the production process (<http://www.jgi.doe.gov/sequencing/cspseqplans2007.html>). Genomic data will augment the largely un-annotated EST sequence data from the Agriculture Canada (c.f. Chapter I, section 1.3.2), and a complete database for the *Tetranychus* genome would be invaluable for rapid identification of orthologs and species-specific sequences for amplification of target genes or putative regulatory DNA regions of interest. A source of chelicerate genomic information would also be useful to provide non-Pancrustacean data-points within molecular phylogenies, with the caveat that relative rates tests be done to identify molecules that do not produce long branch effects as some *Tetranychus urticae* genes have been shown apt to do (e.g. *vasa*).

My colleague A. Khila has reported testing parental RNAi in *Tetranychus urticae*, with only limited success: penetrance rates are apparently quite low, and phenotypes are variable and difficult to interpret even for genes whose phenotypes are expected to be very predictable (e.g. *Tu-Dll* RNAi knockout causes loss of ventral midline integrity and lateral fusion of putative limb primordia, rather than predicted loss of limbs) (Khila 2006b). My own preliminary work on egg injections (c.f. clonal inheritance of dye in holoblastic blastomere divisions, Chapter II) have proved that although single-cell-specific injections are feasible, post-injection environmental conditions allowing survival beyond about 24hrs AEL (e.g. osmotic pressure and humidity), are a challenge to simulate. These points lead me to conclude that *Tetranychus urticae* is not adequately amenable to demonstrations of genetic loss-of-function by either embryonic or parental RNA-interference methods to consider it a 'model' for contemporary evo-devo at present. It may be that RNAi techniques can be improved, or some form of transgenesis established in this species, but if so I still would imagine that the labour-intensive and dexterity-demanding nature of working with spider mites may continue to rule it out as a model organism of choice.

7.3.3 Verdict on *Tetranychus* as a model organism for evo-devo

As previously mentioned, arthropod comparative evo-devo is undergoing a paradigm shift, from measuring trends in gene expression against orthologous gene expression and function in the model insect *Drosophila melanogaster*, to incorporating more and more functional data from disparate taxa.

While I believe that *Tetranychus urticae* is not likely to become tenable for efficient functional studies any time soon, it remains valuable to evo-devo for its phylogenetic position as a derived member of the Acarida, amenable for comparison with other mites (e.g. *Archegozetes longesitosus*), other arachnids (e.g. spiders *Cupiennius salei*, *Achaearanea tepidariorum*, *Pholcus phalangoïdes*, *Steatoda triangulosa*, *Tegenaria saeva*), and basal chelicerates such as Xiphosura (e.g. *Limulus polyphemus*) and scorpions (e.g. *Euscorpius flavicaudis*, *Paruroctonus mesaensis*), as well as other major arthropod or closely related non-arthropod groups. The ability to compare spider mite gene expression profiles with data for the same genes in chelicerates for which gene function assays have become possible, potentially makes inferences more powerful. This rationale seems especially just as those species most impressively amenable to functional study are also members of the Arachnida (e.g. the spiders *Cupiennius salei*, and most notably the spider *Achaearanea tepidariorum*) (Akiyama-Oda and Oda 2006; Schoppmeier and Damen 2001; Stollewerk et al. 2003a). Inevitably however, in such a scenario spider-mite based conclusions will always be subject to criticism, being conditional on final, solid proof of gene interaction and function *in vivo*. Furthermore, spider mites are not only derived within the Chelicerata but are highly specialised, with a number of associated secondary modifications in embryonic patterning (e.g. opisthosomal reduction) and organo-/morpho-genesis (e.g. phytophagous gnathal appendages) that could confound the interpretation of given target gene activity relative to species whose embryos exhibit more ancestral states. In summary, although *Tetranychus* data may be useful to molecular phylogenetics, regarding comparative evo-devo the spider mite would be better considered an interesting choice of specialised arthropod for limited investigation of candidate gene expression, backed up by functional studies in other related chelicerates, rather than being promoted to the intensively-studied status of ‘model’ organism.

7.4 Potential future directions for the study of spider mite limb development and evo-devo

7.4.1 Loose ends in this study of *Tetranychus Dll* regulation

In Chapters III – VI, I have presented and commented on results obtained during my PhD, using a candidate gene approach to assess possible GRN dynamics operating to restrict *Dll* activation to ventral limb primordial *Tu-Dll* transcription foci, detectable in prosomal segments from mid-late germband stages onwards. Due to practical problems and lack of time, several of the steps in characterising gene sequence and expression data for potential antero-posterior and dorso-ventral genetic players in *Tu-Dll* regulation, remain uncompleted. Of particular relevance to a fair assessment of segmental limb primordium specification in spider mites, given more time I would attend as follows to deficiencies in my results:

- i) Detection of *Tetranychus* Distal-less protein activity is desirable in order to confirm actual activity of the transcription factor at the time limb specification is perceived to occur. This would require a spider-mite specific antibody, as G. Panganiban's α -Dll antibody did not work in *Tetranychus* embryos. Given that *Tetranychus* embryo fixation methods may damage proteins, antibody staining may be generally unfeasible in the spider mite and so I would put this 'future aim' at a lower priority than those relating to gene cloning or mRNA detection.
- ii) I would like to be able to thoroughly screen for spider mite *Wnt* genes, in order to characterise which *Wnt* super-family orthologs are present in the genome, and where each may be expressed during development. I would particularly like to know if the hypothetical *Tu-Wnt1* gene has been lost in the spider mite lineage, and look for neo- or sub-functionalisation among *Wnt* paralogs regarding para-segmentation. Theoretically, an effective Wnt screen could be done by screening the *Tetranychus urticae* lambda-Zap cDNA library with a suitable probe, but even more reliable would be to search the complete spider mite genome once it is released by DOE-JGI.
- iii) It would be interesting to finish cloning and sequencing *Tu-en2*, and then detect mRNA transcription for *Tu-en1* and *Tu-en2* paralogs separately: this process merely requires more time using molecular techniques already available. Separate gene expression profiles for *Tu-en1* and *Tu-en2* would provide an indication of which paralog was cloned by the Canadian research group, and an

indication of disparity in transcription domains, hinting at divergent regulation and para-, sub- or neo-functionalisation between paralogs (as discussed also in section 7.2).

iv) Having identified complete coding sequence (and more) for the spider mite *decapentaplegic* ortholog *Tu-dpp*, it was bizarre that I could not detect its transcription by *in situ* mRNA hybridisation. Colleagues in Canada also report that they could not detect *Tu-dpp* gene expression in whole mount embryos, and so perhaps it would be worth synthesising a *Tetranychus*-specific Dpp antibody to detect ligand localisation instead. I would like to check the *Tetranychus urticae* genome when it is published to confirm my *Tu-dpp* sequence, possibly examine upstream *cis*-regulatory regions and decide whether to abandon further problem-solving investigations or not.

v) It would be interesting to find or synthesise a cross-reactive dpMAPK antibody to augment *TER-a* mRNA transcription patterns. Similarly, cloning of a spider mite *EGF* ligand ortholog could provide an indication of how far secreted EGF ligand molecules travel before EGF-R and MAPK activation, and the degree to which auto-regulation may be occurring between EGFR-mediated MAPK signalling and *EGF* transcriptional activation.

vi) Regarding *Bithorax* complex *Hox* genes, it would be ideal to take the opportunity afforded by a complete genome database to confirm that the hypothetical *Tu-abdA* has been lost from the spider mite lineage. This confirmation would lend more credibility to the proposition that apparent *abdominal-A* gene loss (c.f. Chapter VI) is real and may be correlated with the complete reduction of the spider mite opisthosoma. In addition, it may be informative to synthesise a *Tetranychus*-specific Ubx antibody as the commonly used FP6.87 UbdA antibody does not cross-react in *Tetranychus* due to unusual epitope divergence, and yet *Tu-Ubx* protein detection could allow fruitful comparison with mRNA data, especially regarding the nature of any post-transcriptional *Tu-Ubx* regulatory control.

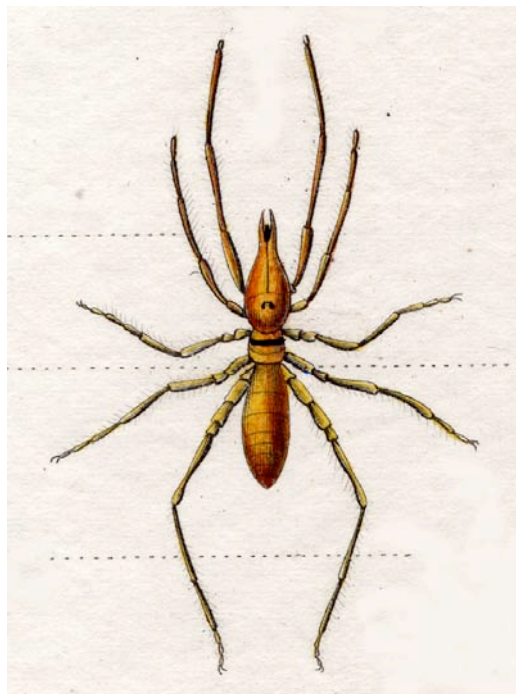
vii) Having identified spider mite orthologs of both *Sp8/9* and *Sp1/5/4* sub-family genes I would be interested to extend the gene fragments available such that mRNA transcription during embryogenesis could be detected *in situ* with RNA probes. The aim of this would be to test whether or not *Sp8/9* gene expression in particular is linked with any stage of limb development, and hence to expand the model based on insect and vertebrate *Sp* gene activity, that *Sp* genes may play deeply conserved roles in initiating and/or mediating differential proximo-distal limb outgrowth.

7.4.2 Last word: The place of *Tetranychus* in the future of evo-devo

As mentioned previously, *Tetranychus urticae* embryogenesis is not currently amenable to gene function assays (section 7.3), but it is nevertheless valuable to obtain gene expression data for as many arthropod species as possible, including the spider mite, to reveal common trends or innovations in temporal or spatial gene or protein regulation. Awareness of spatio-temporal gene expression patterns across as wide a taxonomic range as possible can fuel hypotheses about the nature of GRNs operating in key patterning processes, that can then guide the focus of functional studies in corner-stone species for which gene function *is* clearly demonstrable. Hence, in the future I would choose to continue to look at gene expression profiles in *Tetranychus urticae*, aiming to obtain a complete impression of spatio-temporal dynamics for candidate genes implicated as critical to common arthropod processes such as segmentation, limb specification and neural patterning. It is fair to say that the practical difficulties encountered when handling this troublesome animal mean that I would recommend using it for occasional, speculative studies, while channelling primary resources to more reliable and workable species. If the vision for *Tetranychus urticae* evo-devo can be fruitfully limited to speculative exploration aiming to broaden the phylogenetic base of genetic information with which to form new and refine old hypotheses, it could remain a small, quirky but valued player on the stage of comparative evolutionary developmental biology.

CHAPTER VIII

MATERIALS AND METHODS



Faux scorpion - Galéode Aranéide

8.1 *Tetranychus urticae* embryo preparation

8.1.1 Spider mite husbandry

I cultured spider mites of species *Tetranychus urticae* in a vivarium, where they feed and lay eggs on young French bean plants of Prince/Tendergreen varieties. The culture was maintained at room temperature and exposed to temperate light conditions.

8.1.2 Spider mite embryo collection

Collection of embryos for DNA or RNA extraction

I made large embryo collections prior to DNA/RNA isolation, flash freezing collected eggs in liquid nitrogen and storing at minus 80°C when several samples were required to make up an adequate mass for nucleic acid extraction. Heavily mite-infected bean leaves were removed from plants and rinsed for 2' in 0.05% Tween-20 detergent, with gentle stirring. Adults and nymphs were separated from eggs by rinsing through a Sigma 200µm-mesh sieve, and eggs were subsequently removed from solution by rinsing through a 70µm-mesh nylon cell culture basket.

Collection of embryos for fixation

Prior to embryo fixation and whole mount detection of specific mRNA transcripts or protein activity (c.f. sections 8.1.3 and 8.1.4), I found very small collections of spider mite eggs to be sufficient. I rinsed just two or three infected bean plant leaves in 0.05% Tw-20 and filtered eggs out using a 70µm nylon mesh basket as before, then immediately proceeded with fixation.

Collection for timed embryo fixation

Developmental timing studies required selection and incubation of individual live embryos at specific temperatures, and hence a different means of collection. I removed and examined single infected leaves under a dissection microscope to locate embryos of the desired stage, using transverse top highlight cytoplasmic haloes surrounding cleavage nuclei. I gently lifted embryos from the leaves using a few hairs from a paintbrush, and transferred them to the underside of a small piece of uninfected leaf. Once 20-30 embryos were collected, the piece of leaf on a moist disc of filter paper in a lidded Petri dish, was incubated at a specific temperature (either 4°C, 18°C or 25°C) and for a specific duration. Prior to any fixation or staining, I removed embryos one at a time from the leaf surface, using the slightly bent tip of a fine syringe needle (G30) to gently spoon eggs into a small volume of PBS (e.g. 100µl in a 0.5ml eppendorf tube).

Reagents Rinse: 0.05% Tween-20

 PBS: 1x solution

8.1.3 Fixation of spider mite embryos for *in situ* hybridisation

I developed a novel protocol to fix embryos for *in situ* hybridisation, as the protocols obtained from other *Tetranychus urticae* researchers (Dearden et al.) provided unacceptably poor results or very low % success regarding subsequent mRNA/protein detection: The protocol minimised the duration, and hence RNA-damaging effects, of exposure to boiling temperatures. I found a high-temperature mild fixation to be necessary to make egg chorion and vitelline membranes amenable to manual removal prior to effective formaldehyde fixation.

Pre-in situ hybridisation fixation protocol

1. Immediately after collection, using a Pasteur pipette and rinsing with 1x PBS, transfer embryos from 70µm-mesh basket to 1.5ml eppendorf tube. Allow 5' for embryos to settle.
2. Remove as much PBS as possible, then boil for 45" (e.g. by submerging the lower half of the tube into the water-filled cell of a 100°C hot block).
3. Remove from heat, immediately adding 600µl chilled 1x PBS and a little ice. Shake vigorously for 10", then store the embryo tube on ice for remainder of fixation process.
4. Transfer embryos to work within a droplet of PBS on a plate of optically clear silicone elastomer (Sylgard® 184, Dow Corning). Dissect membranes off using one G30.5 (0.35mm gauge diameter) BD Microlance tapered syringe needle, and one G30.5 needle into which has been mounted (with melted bee's wax) 25µm diameter tungsten (W) wire, sharpened to a needle-point by electrolytic degradation in NaOH. Roll/push embryos in opposition to basal surfaces and then hold by any initial tear in the membranes as the tear is enlarged with the W needle until the embryo is completely free. Transfer dissected eggs with a 10µl tip to a droplet of PBS on a second, chilled Sylgard plate. Continue for up to 50', being careful to replace the lid of the Sylgard collection plate between each addition of more embryos, to protect from dust and other particulates.
5. Remove PBS from around the embryos gently with G30.5 syringe. Replace with 100µl 4% formaldehyde in 1x PBS. Allow embryos to fix for 45-60' on a gently rotating shaker at room temperature (RT).
6. Rinse 1x with PBS (remove droplet of fix, replace with droplet of PBS), then 2x with PTw. Wash in PTw for 10', on a gently rotating shaker.
7. Transfer embryos to pre-hybridisation buffer, either directly into a glass staining well for immediate use or into storage at -20°C.

Reagents

PBS:	1x solution
PTw:	1x PBS + 0.05% Tween-20
NaOH:	5M solution

8.1.4 Fixation of spider mite embryos for antibody staining

To balance the need for both a degree of membrane fixation and to preserve the secondary structure of epitopes recognised by specific antibodies, the following protocol was developed, with slightly reduced boiling time and shorter gap between membrane removal and fixation:

Pre-antibody staining fixation protocol

1. As in the *in situ* fixation protocol, transfer embryos from the 70µm-mesh collection basket to ~500µl 1x PBS in a 1.5ml eppendorf tube. Allow 5' for embryos to settle.
2. Remove as much PBS as possible, then boil for 30"- 35", placing the open eppendorf in the water-filled cell of a 100°C hot block.
3. Remove from the hot block, immediately adding 600µl chilled 1x PBS and a little ice. Shake vigorously for 5-10", then store the tube of embryos on ice.
4. Dissect membranes from around the yolk and embryonic tissue as in the *in situ* fixation protocol, using sharpened G30.5 + W(tungsten) needles and working within a droplet of PBS on a Sylgard plate. Transfer dissected eggs to a 100µl droplet of PBS on a chilled Sylgard plate, continuing for up to 35' and always replacing the lid of the Sylgard plate to protect dissected embryos from dust/particulates.
5. Remove PBS from around the embryos gently with G30.5 syringe. Replace with 100µl 4% formaldehyde in 1x PBS. Allow embryos to fix for 15-25' on a gently rotating shaker at RT.
6. Rinse 1x with PBS (i.e. remove droplet of fix, replace with droplet of PBS), then 2x with PTw. Wash in PTw for 10', on gently rotating shaker.
7. Transfer embryos to either directly to 100µl PTw in a glass staining well for immediate use, or dehydrate through a methanol series for storage at -20°C. Dehydration increments are 0% -> 30% -> 50% -> 75% -> 100% methanol (in PTw), and embryos are rehydrated through similarly increasing concentrations of PTw prior to staining.

Reagents

PBS:	1x solution
PTw:	1x PBS + 0.05% Tween-20

8.2 Molecular Cloning and Sequencing

8.2.1 Degenerate PCR

The standard polymerase chain reaction (PCR) provides a powerful means of amplifying, many millions of times, a highly specific target stretch of DNA (termed the 'amplicon') from within a longer, double-stranded genomic or cDNA sample.

Basic PCR

The procedure for basic PCR was originally developed by Saiki et al (1985), and is based upon three phases of activity: (i) denaturation and strand separation of DNA template at $>94^{\circ}\text{C}$, (ii) annealing of specific oligonucleotide primers to 5' and 3' ends of a target DNA sequence at $40\text{--}60^{\circ}\text{C}$ and (iii) extension of the target/amplicon from sites of primer binding at 72°C , catalysed by a thermostable Taq polymerase (derived from the thermophilic bacterium *Thermus aquaticus*) in a solution of Mg^{2+} -buffered nucleotide bases (deoxynucleotide triphosphates; dNTPs). Successful amplification of target DNA fragments depends largely on optimization of primer design, thermal cycling conditions in the PCR program and Mg^{2+} concentration. Increased $[\text{Mg}^{2+}]$ causes greater ionic interference, promoting an 'open' structure and lowering T_m DNA. Further details are given in: *Promega Protocols and Applications Guide* (2005), 'Chapter 1: Nucleic Acid Amplification', and in *PCR 2* (1995), ed.s M. McPherson, B. Hames and G. Taylor, 'The Practical Approach Series', IRL Press.

Degenerate PCR

Degenerate PCR is a method of isolating a gene *de novo* from one species, with the aid of sequence data previously obtained for its orthologs in others. Alignments reveal domains of high sequence conservation between the known taxa, and primers are targeted against such regions. Primers incorporate a specific 'core' region allowing no codon base variation, and a 'clamp' region in which Universal nucleotide molecules can bind to multiple possible nucleotides (hence the name 'degenerate'), albeit with reduced affinity relative to non-degenerate bonds. Details of the Universal Code are available in Alberts et al. *The Cell* (1994) (Alberts et al. 1994). Degree of degeneracy ($^{\circ}\text{D}$) is a factor of the numbers of nucleotides bound by each non-specific base; for a primer with N (=A/T/C/G) and R (=C/G), $^{\circ}\text{D}$ is $N=4 \times R=2 \therefore ^{\circ}\text{D}=8$. Values of $^{\circ}\text{D}$ are minimised (ideally <64) to maximise the probability of successful primer binding to homologous gene sequences. To optimise annealing of degenerate primers to their target unknown sequences, $[\text{Mg}^{2+}]$ and PCR cycle conditions may be modified. In 'touchdown' PCR, annealing temperatures are gradually decreased for initial higher-specificity binding and reduced likelihood of amplifying erroneous sequence material during preliminary PCR cycles. If screening for multiple related genes with shared motifs,

early annealing temperatures may be increased gradually for maximal binding to the shared domain and amplification of as many gene family members as possible. This second approach was used, for example, by deRosa et al. (1999) in an extensive Hox screen in species from the major metazoan phyla (de Rosa et al. 1999). I applied degenerate PCR to amplify *Tetranychus Dll*-regulatory gene orthologs from either genomic DNA, total RNA (obtained using QiaGen DNA/RNA extraction kit) or a lambda-Zap embryonic cDNA library (obtained from Agriculture Canada). Distinct PCR programs and conditions were developed to facilitate amplification of the various candidate gene homologs. The BioRad iCycler and ThermoHybaid Px2 PCR machines were used, and the details of both standard and specialised PCR programs and magnesium concentrations used are summarised in Table 8.2.1 below. Details for degenerate primers that I used to amplify preliminary fragments of putative *Tetranychus Dll*-regulatory gene orthologs are given in Appendix 2.

PROGRAM NAME	PCR STEPS (PLUS # CYCLES)
Standard 50	5' @ 96°C denaturation 30" @ 96°C, 30" @ 50°C annealing, 1' @ 72°C extension (x40) 7' @ 72°C final extension 4°C ∞
Standard 48	5' @ 96°C 30" @ 96°C, 30" @ 48°C, 1' @ 72°C (x40) 7' @ 72°C 4°C ∞
Deg-Up 6x	5' @ 95°C 30" @ 94°C, 30" @ 46°C (↑1°C/cycle), 1' @ 72°C (x6) 30" @ 94°C, 30" @ 50°C, 1' @ 72°C (x40) 5' @ 72°C 4°C ∞
Deg-Up 4x	5' @ 95°C 30" @ 94°C, 30" @ 46°C (↑1°C/cycle), 1' @ 72°C (x4) 30" @ 94°C, 30" @ 50°C, 1' @ 72°C (x40) 5' @ 72°C 4°C ∞
Deg-Down 6x	5' @ 95°C 30" @ 94°C, 30" @ 56°C (↓1°C/cycle), 1' @ 72°C (x6) 30" @ 94°C, 30" @ 50°C, 1' @ 72°C (x35) 5' @ 72°C 4°C ∞
Deg-Down 10x	5' @ 95°C 30" @ 94°C, 30" @ 60°C (↓1°C/cycle), 1' @ 72°C (x10) 30" @ 94°C, 30" @ 50°C, 1' @ 72°C (x35) 5' @ 72°C 4°C ∞

Deg-Down 48	5' @ 95°C 30" @ 94°C, 30" @ 54°C (↓1°C/cycle), 1' @ 72°C (x6) 30" @ 94°C, 30" @ 48°C, 1' @ 72°C (x35) 5' @ 72°C 4°C ∞
Gradient 48-54	5' @ 96°C; 30" @ 96°C, 30" @ 54-48°C, 1' @ 72°C (x30); 7' @ 72°C; 4°C ∞
Gradient 50-58	5' @ 96°C; 30" @ 96°C, 30" @ 58-50°C, 1' @ 72°C (x30); 7' @ 72°C; 4°C ∞
Deg-Dll	5' @ 96°C 30" @ 94°C, 30" @ 43°C, 1' @ 72°C (x5) 30" @ 94°C, 30" @ 50°C, 1' @ 72°C (x30) 7' @ 72°C 4°C ∞

Table 8.2.1 PCR program details. **N.B.** Standard 50 program is used for colony PCR (section 8.2.5) , and further PCR sub-programs specific to iPCR or RT-PCR protocols are given within the text where relevant.

8.2.2 Reverse Transcriptase PCR

Reverse transcriptase PCR (RT-PCR) enables amplification of target gene sequence using total RNA or purified mRNA as a template from which single-stranded, then double-stranded DNA is transcribed. The whole procedure, including RNA isolation, is carried out in an RNase-free environment. I extracted total RNA from *Tetranychus* egg material using a QIAquick RNA isolation kit; see kit handbook for reagent details etc. Boehringer Expand RT was used as the reverse transcriptase for first-strand synthesis, and second strand cDNA synthesis was carried out with Roche Taq polymerase II and standard PCR conditions. Relevant protocol details are given below:

i) First strand synthesis (FSS) protocol

- To 0.2ml tube add: 1µl RNA (1µg/µl total RNA)
5µl DEPC H₂O =6µl total
- Denature RNA at 65°C for 10' then transfer 6µl reactions to ice
- Add 13.5µl aliquot master mix I: 4µl 5x buffer
2µl DTT
2µl dNTPs (in DEPC H₂O) 5mM
0.5µl RNase Inhibitor
3.5µl DEPC H₂O
1µl Primer = Hexanucleotides (1:10)
0.5µl Expand RT =13.5µl mix + 6µl RNA = 20µl total
- Run FSS PCR program (PCR products stable at 4°C few hrs or -20°C overnight)

25°C 10'	(annealing)
42°C 1hr	(reverse transcription)
95°C 10'	(inactivation)
4°C ∞	

ii) *Second strand synthesis (SSS) protocol*

- Add a 30µl aliquot of Master mix II to each 20µl FSS reaction

Master mix II: 5µl 10xbuffer

0.5µl 100pmol Fwd primer

0.5µl 100pmol Rev primer

2µl dNTP 5mM

22.35µl dH₂O

0.25µl Taq Pol II

=30µl mix + 20µl FSS = 50µl total

- Amplify second-strand DNA using a standard PCR program
- Run 3µl of the 50µl final SSS PCR product on 1% agarose gel to verify amplification of fragment(s) within the correct expected size range

8.2.3 Inverse PCR

Inverse PCR (iPCR) is based on cutting genomic DNA with a known restriction enzyme, re-ligating it to form circular DNA, and amplifying by PCR outwards into a region of unknown DNA sequence from specific primers (nested pairs) facing outwards 5' and 3' of a short, previously cloned sequence within a DNA region that has become circularised. A more detailed account of iPCR is available in Erlich (1985) (Erlich 1985). The iPCR protocol used in this project was adapted from Cook et al. (2001), and is schematically represented in **Figure 8.2.3** on the next page (Cook et al. 2001). See Appendix 2 for non-degenerate iPCR primers used to amplify *Ubx*, *EGFR* and *∂pp*, and those designed for putative '*abd-A*' and '*TER-1*' orthologs.

8.2.4 cDNA library screening

The *Tetranychus urticae* Lambda-ZAP-CMV-XR cDNA library was made from embryos at mixed stages of development from 0-48hrs AEL, and donated by the Southern Crop Protection and Research Centre in Ontario, Canada. The cDNA library was titred (28,000pfu/µl) and screened according to the protocol, and using the reagents outline in, the Stratagene Lambda-ZAP-CMV-XR

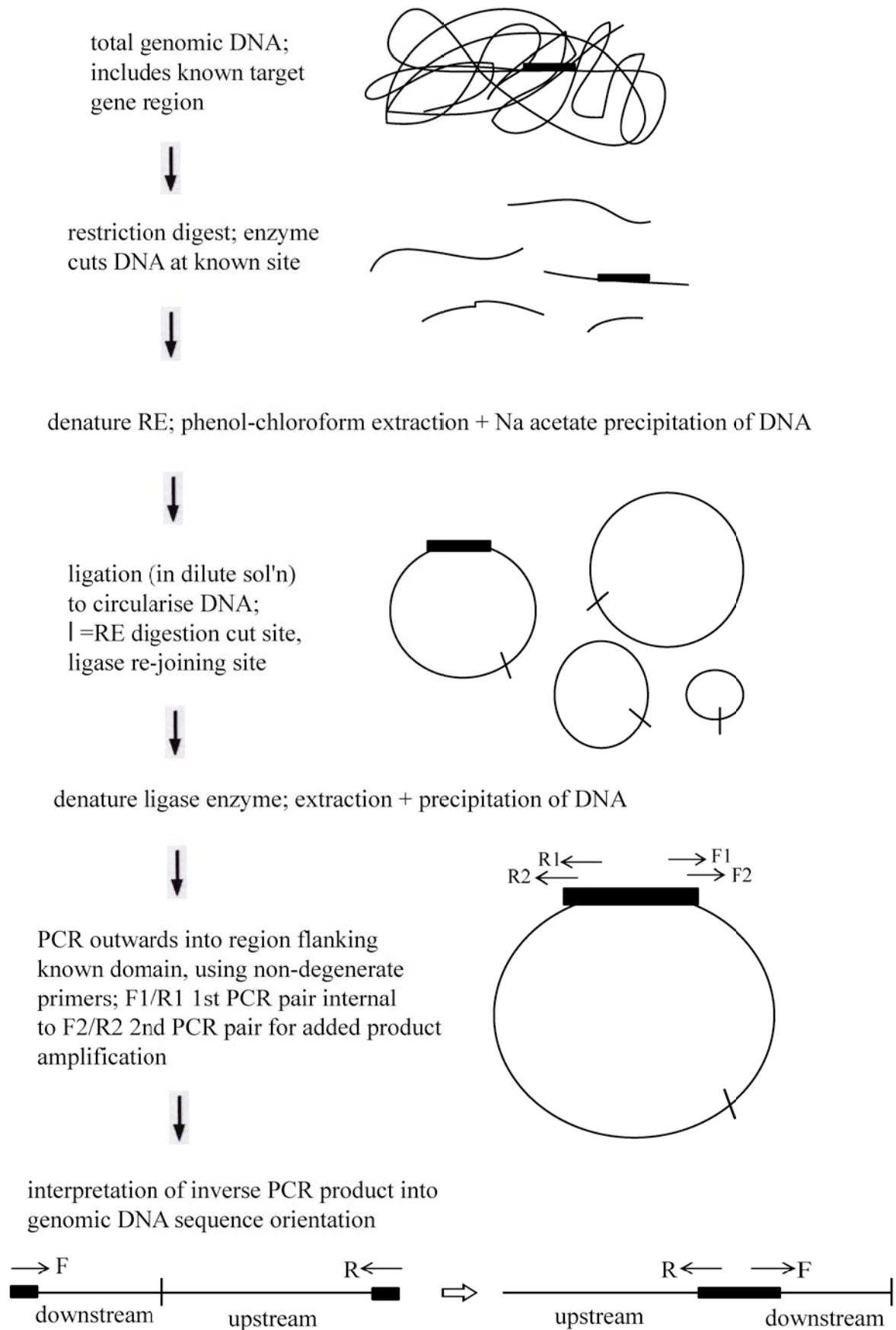


Figure 8.2.3 Schematic diagram of inverse PCR protocol, after Averof (1994)

instruction manual. I excised target gene inserts from pCMV-Script EX vector, also according to instructions in the Stratagene Lambda-ZAP-CMV-XR manual. I re-amplified and purified screening probe DNA from 200-350bp long fragments amplified during prior degenerate PCR (see Appendix 2 for primers used). I labelled probes with horseradish peroxidase (HRP) using the ECL™ direct nucleic acid labelling and detection system (Amersham), and successful hybridisation to target gene-containing plaques was detected by chemi-luminescence on Amersham ECL Hyperfilm™. Details of protocols for probe labelling and detection can be found in the ECL™ kit manual.

8.2.5 Cloning

I obtained purified target gene fragments by PCR purification or gel extraction using QIAquick® kits, according to protocols in the QIAquick manual. I cloned DNA fragments by insertion into Promega pGEM®-T Easy vector as described in Promega Technical Manual No. 042: pGEM®-T and pGEM®-T Easy Vector Systems. I used a standard PCR program (c.f. Table 8.2.1) and standard ionic conditions for colony PCR, with vector primers T7 + SP6 or MB40 + Rev48 at 25pmol. I prepared purified plasmid-insert DNA using the QIAprep® Miniprep kit, from 2ml 1:1000 CARB cultures incubated at 37°C overnight, also taking a 200µl bacterial culture sample to preserve plasmids in glycerol stock at minus 80°C.

8.2.6 Gene sequencing

I carried out sequencing reactions as follows:	2.5µl purified plasmid DNA template
	2µl 2.5x buffer
	1µl T7/SP6 vector primer (5pmol)
	2.5µl ddH ₂ O
	<u>2µl BigDye® sequencing mix</u>
	=10µl total

The sequencing PCR program included: 3' @ 96°C initial denaturation followed by 25 cycles of 20" @ 96°C, 10" @ 50°C, 4' @ 60°C, before holding for ∞ @ 4°C. I cleaned up sequencing reactions, removing unincorporated low molecular weight components such as primers, as follows:

Sequencing clean-up protocol

1. Add 500µl autoclaved G50 Sephadex beads at 6.6g/100ml ddH₂O to spin columns in 1.5ml eppendorf tubes
2. Spin for 1' at low speed (e.g. ~5000rpm)
3. Transfer spin columns to new 1.5ml tubes; apply 10µl sequencing reaction to tip of Sephadex bead column
4. Spin for 1' at low speed (e.g. ~5000rpm)
5. Purified product is in the eluate; dry on Speedy Vac for ~30' at 50°C and store at minus 20°C until chromatograms are ready to be run

Chromatograms were run at the sequencing facility of the Natural History Museum, London or the Department of Genetics, Cambridge. Resultant sequence files were analysed using Sequencher version 4.5. I discarded any bad sequence read-out results based on regularity and readability of chromatogram data, and identified restriction enzyme (RE) cut sites within the program, used for example to assemble iPCR products in correct genomic orientation and/or to remove unwanted vector sequence. Multiple alignments of identical or contiguous cloning products ('contigs'), were created and edited where necessary prior to generating consensus sequence data that could be exported for further sequence analysis and identification.

8.3 Phylogenetic analysis of putative orthologs

8.3.1 Molecular phylogenetics theory

Multiple sequence alignments (section 8.3.3) represent the raw material of datasets subjected to phylogenetic analysis, aiming to examine and represent relatedness within a group of genes or proteins. In this project, homologous characters within each sequence, i.e. those occupying equivalent positions in the alignment, were compared between all taxa according to Bayesian inference search criteria (section 8.3.4) or pair-wise maximum likelihood (ML) measurements (section 8.3.5). The output of any such analysis is a **phylogram**, a graphic depiction of the degree of relatedness between

the nucleotide or amino acid sequences analysed. Sequences of known lower relatedness than those containing the gene or genes whose relationships are in question is included in the dataset as an **outgroup**. Remaining sequences form a specific branching pattern away from the outgroup 'root', and the statistical probability supporting each inferred relationship is given numerically and in the branch lengths which represent the relative rate of molecular substitution for each sequence in question. It is in visualising organisation of branching, or **tree topology**, that the homology of a putative orthologous gene can be confirmed or refuted, dependent on inclusion or exclusion from a clade of orthologs (**ingroup**) from other taxa.

Confounding problems in obtaining a 'true' topology representing the relationship of various sequences to a hypothetical ortholog are:

1. Insufficient **phylogenetic signal**, and
2. **Long branch effects**.

Regarding **problem 1**, if a short sequence is compared, or one in which a heavy proportion of amino acids are conserved, there may be insufficient differences between taxa to determine the molecular evolutionary relationships between them. Regarding **problem 2**, if any sequence within the dataset evolves at a significantly faster rate (relative rate) than the others, it may be artificially drawn towards the base of the topology and/or disrupt the branching pattern of outgroup and ingroup sequences. Although taxa causing long branch effects can be assessed by **relative rates** tests, and subsequently eliminated from analyses, both potential problems 1. and 2. may confound effective confirmation (or refutation) of orthology status for putative homologs under investigation.

8.3.2 Preliminary gene identification

Orthology of cloned genes to candidate *Dll*-regulatory genes was initially demonstrated by comparing sequence data with that available for other species in GenBank. GenBank accessions can be searched in comparison with a candidate sequence through the BLAST website(Altschul et al. 1990). Nucleotide and polypeptide sequence identity were checked using BLASTn and tBLASTx, searching against nucleotide and translated database sequences respectively (Altschul et al. 1990).

8.3.3 Generation of multiple nucleotide/amino acid alignments

Alignments of target gene orthologs from many species were created by downloading FASTA format files of coding nucleotides or amino acid sequence data from GenBank at the NCBI website. Prior to

alignment of sequences FASTA entries must be saved in Text Only format, arranged as below for correct processing:

```
>species1_gene ↵  
nucleotide sequence 1 in frame, or amino acid sequence 1 (no gaps)  
>species2_gene ↵  
nucleotide sequence 2 in frame, or amino acid sequence 2    etc...
```

I aligned nucleotide sequences using the program TranslatorX3 (operated *via* Terminal interface), and aligned amino acid sequences using Clustal-X (Jeanmougin et al. 1998). To convert coding nucleotide sequences into amino acids, or to ensure that a given sequence will be read by subsequent programs in reading frame 1, I used the program TranSeq. TranSeq is online at www.ebi.ac.uk/emboss/transeq/index.html, providing a service to translate any nucleotide sequence in forward and/or reverse directions, and in all 3 frames so that the correct ‘coding frame’ can be identified for any problem sequence.

Before alignment within the program TranslatorX3, FASTA multiple sequence files must be opened in MacClade and exported in NBRF format (Maddison and Maddison 2000). TranslatorX3 aligns in-frame coding nucleotides, processing an inferred sequence of translated codons to create an alignment within either the program Clustal-X, Muscle or T-Coffee (Jeanmougin et al. 1998). Commands to run files within TranslatorX3 via the Terminal interface are in **Appendix 4.1**. Output files are multiple nucleotide FASTA sequences, aligned and requiring conversion from Unix to Macintosh in a text program, e.g. BBEdit Lite 6.1 or TextWrangler.

8.3.4 Defining Exclusion and Inclusion Sets

In MacClade (Maddison and Maddison 2000) I selected regions of the alignment to be included in subsequent sequence analysis - i.e. characters deemed homologous and phylogenetically useful. Uninformative characters comprise a dataset for exclusion (‘ExSet’), in contrast to a complementary dataset (‘InSet’) to be included in further Bayesian or Maximum Likelihood analysis. See **Appendix 4.2** for protocol and file formatting details relevant to defining InSet and ExSet characters within MacClade. Of note, modified alignments must be saved as NBRF format to preserve compatibility with subsequent phylogeny programs.

8.3.5 Bayesian analysis

NEXUS files for nucleotide alignments were subjected to Bayesian inference analysis, using MrBayes phylogenetic inference package version 3.1.1 (Ronquist and Huesenbeck 2003; Ronquist and Huesenbeck 2005). A 'Bayes Block' is written beneath the alignment matrix as described in **Appendix 3.4**, defining a set of instructions about settings and characters for the program to include. Visual inspection of Log likelihood (LnL) values permitted assessment of when the 'best', most probable representation of relationships had been stably determined for the sequences analysed, after which the resultant 'consensus tree' topology is exported for processing.

Analysis settings

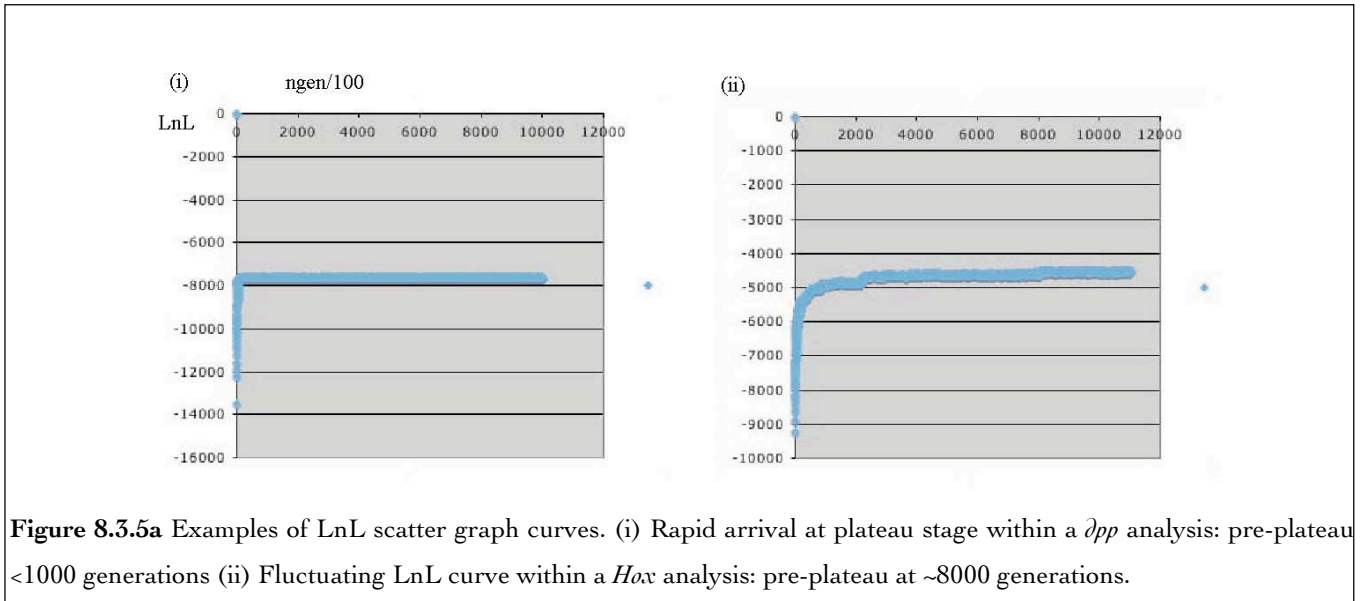
- i) Third codon positions ('3rdposn') are excluded due to high functional lability at the third base of any given codon and likely associated mutational saturation as multiple transformations and reversions occur, masking useful phylogenetic information.
- ii) Exclusion set values, given in the nexus text file, are present as 'EXSET' values nested within BEGIN SET/BEGIN ASSUMPTIONS of the Bayes Block.
- iii) Number of states ('nst') refers to complexity of states or assumptions included in the phylogenetic model. For example, allowing two states would treat transitions and transversions as being of unequal probability and evolutionary significance relative to the nature of the 'ancestral' base. I prescribed 'nst = 6', considering that 6 states can distinguish more subtle factors affecting the probability of a given base changing/reverting to any other.
- iv) 'ngen' refers to number of generations, in the form of searches made for the 'best tree', and 'samplefreq' (sample frequency) refers to the interval between recordings, or frequency at which Log Ln values for trees are recorded whilst the analysis is running. Typically, sampling every 100 trees, and running 700,000 – 1,000,000 generations proved adequate for my purposes.

Assessing stable tree topology

Once the number of specified generations have been analysed, various output files are created, relevant ones being suffixed '.t' and '.p' (duplicate versions 'run1.p/t' and 'run2.p/t' exist for each). Files suffixed .t (tree) contain tree topology information, and .p (probability) files most notably the log likelihood (LnL) data for each tree permutation recorded every 100 generations in the search for the 'best tree'. Additional data in the .p files records probability values for each of the 6 'rates'/types of base change specified previously; e.g. r(C->G).

To decide whether further analysis was required, or if log likelihood values were stable enough to generate a final 'consensus tree' for the dataset, I opened either one of the duplicate .p files ('filename.nex.run1.p' or 'filename.nex.run2.p') in Excel and examined an XY scatter graph of LnL values for *all* generations, made in ChartWizard. A plateau in LnL values implies that a reliable

consensus tree has been reached within the population of possible ‘best’ trees (see Figure 8.3.5a), and the point at which an apparent plateau is *first* reached - i.e. the generational extent of the ‘pre-plateau region’ is recorded, as trees before this point are unstable and therefore discounted from subsequent consensus tree calculations.



To determine whether the apparent plateau represented genuine arrival at consensus tree values, or alternatively if a slow increase (and hence improvement) in LnL values was still occurring, I generated an XY scatter graph from the most recent 300-500 LnL values recorded. The presence of an overall **horizontal trend** indicated stable tree topology values, required to stop the analysis and progress to consensus tree calculation. There may or may not be fluctuations around the mean within the plateau, as in the two examples in Figure 8.3.5b. If the values were found to be increasing, however, I carried out further analysis, defining the number of extra generations (e.g. 100,000) within Terminal and executing MrBayes as before.

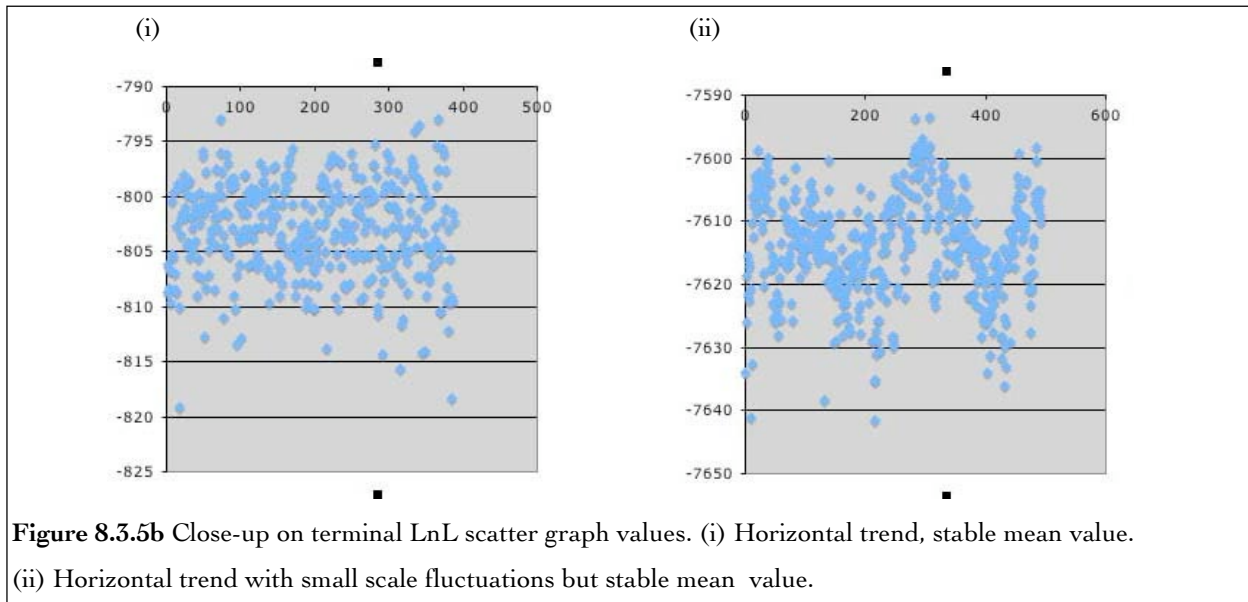


Figure 8.3.5b Close-up on terminal LnL scatter graph values. (i) Horizontal trend, stable mean value. (ii) Horizontal trend with small scale fluctuations but stable mean value.

Consensus tree building

Having confirmed that the analysis had converged on stable LnL values, I created a consensus tree by summing data for the population of trees in the plateau region. Data for trees in the pre-plateau region was discarded as mentioned before, due to instability and sub-optimal inference support for the relationships posited in those trees. Instructions required within the MrBayes Terminal command shell, were:

```
sumt filename = file.nex burnin= #
```

where 'file.nex' refers to the original NEXUS file name and # refers to the number of generations in the pre-plateau zone - i.e. those values to be excluded by the program when calculating a consensus tree. Consensus tree files are created using a 50% majority rule in which the program retains the 50% most abundant trees in the population, disregards the rest, and hence 'evolves' the population towards the most favoured tree topology. Posterior probabilities are given as support values from 0.00 (no support) to 1.00 (100% support) for each branch. Posterior probabilities are not as strict as Bootstrap values, with probabilities >0.90 required to indicate genuinely good support.

Viewing consensus trees

Consensus tree data is held in an output file with '.con' suffix. Having converted the .con file from Unix to Mac in BBEdit or TextWrangler, I viewed trees in TreeView PPC Version 1.6.6 (Page 2001). Representations of trees in TreeView could include posterior probability scores at nodes, and available tree types include radial, slanted/rectangular cladograms or, as was most often the clearest

mode of displaying information, a **phylogram**: i.e. a rectangular cladogram including branch length information. Consensus trees may be rooted with a specific taxon using options available in TreeView: a specific outgroup²⁵ may be defined by choosing one species from among the aligned taxa, transferring it to an 'outgroup box' and re-rooting the tree automatically once this new outgroup is applied.

Dealing with long branch taxa

The presence of any anomalously long branch- or otherwise nonsensical taxa, potentially disrupts the integrity, resolution and significance of relationships represented in any tree. Such 'problem' taxa diminish the power of the analysis to locate the 'best' tree, destabilising LnL values and reducing absolute likelihood measures within all sets of relationship compared. Having identified problem taxa within any preliminary consensus tree, I repeated the whole Bayesian analysis after removing such taxa from the sequence dataset in one of two possible ways as described below:

- (i) One line above the command 'exclude...' in the Bayes Block, add the command 'delete', followed by the species name(s) to be excluded, each name separated by one space and written exactly as in the alignment. Save .nex file with a new name to prevent appending new consensus results to the previous analysis, and re-execute complete analysis.
- (ii) Open original .nex file in MacClade and delete long branch sequences manually. Save .nex file with a new name and re-format the Bayes Block as required. Execute a complete analysis of the reduced dataset, confirming stable LnL values etc. as previously.

I stored images of final consensus trees, opened and rooted in TreeView, by either (i) saving and printing to .pdf, or (ii) saving as a Graphics File, exporting a '.pict' file. For any further editing or annotation, both types of image file (.pdf/.pict) were readable in applications such as Adobe Photoshop or Illustrator.

²⁵ It is also possible to define the outgroup before a .con file is generated. Returning to the active Terminal shell directly after demonstrating a stable LnL plateau in Excel, the command 'outgroup' is followed with the name of the desired species *exactly* as it appears in the alignment (e.g. outgroup BranchiostomaDlx). This command line is immediately followed by the command to sum all the tree data (e.g. sumt filename = DllDlx.nex burnin = 1000) so that outgroup requirements are incorporated into consensus tree calculations.

8.3.6 Maximum Likelihood analysis by Quartet Puzzling

NEXUS files for amino acid alignments were subjected to Maximum Likelihood analysis by quartet puzzling using the program TreePuzzle Version 5.0 (Schmidt et al. 1999). Quartet puzzling was only used during the earlier phases of this PhD, as MrBayes became a more favourable option later on. This was due partly to availability but mainly to the ability of a Bayesian analytical model to consider codon bias as well as a combination of other nucleotide/amino acid state factors, increasing the likelihood of generating tree topologies of superior support (Ronquist and Huesenbeck 2003; Ronquist and Huesenbeck 2005). The computational method used to execute analyses in TreePuzzle 5.0 was as follows:

1. Open and execute .nxs file in PAUP version 4.0-β (Swofford 1998).
2. Within Data; Include/Exclude Characters; exclude characters from the ExSet defined in MacClade alignment.
3. Follow 'Export Data' commands to save InSet sequences as a .dat file in PHYLIP format. Within the .dat file (Figure 8.3.6a), the top left number refers to the number of taxa compared, the top right number is the number of characters compared and the outgroup for the resulting tree is the species at the top of the taxon list.

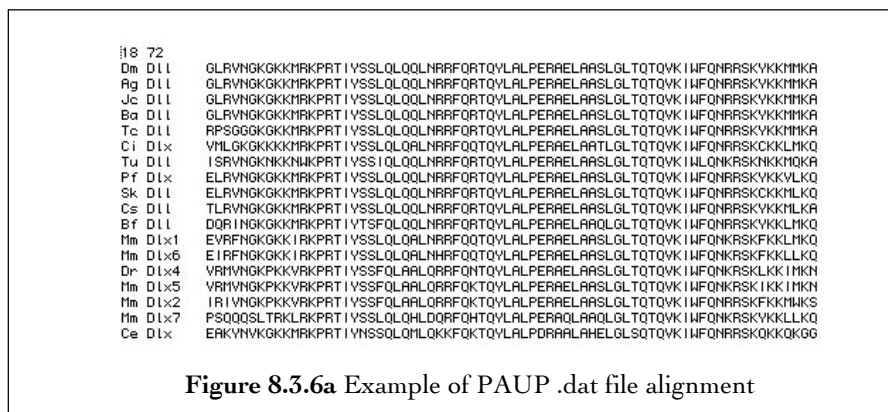


Figure 8.3.6a Example of PAUP .dat file alignment

4. Open TreePuzzle-5.0 folder; Transfer .dat file, renamed as 'infile.dat'.
5. Select TreePuzzle file to start the program, setting optional criteria as follows:
 - Puzzling steps: Maximum limit of 10,000 for theoretical optimisation of the resultant tree; in general I found 1000 puzzling steps to be adequate for stable optimised tree topology
 - Substitution model: WG (Whelan-Goldman)
 - Outgroup: The taxon at the top of the .dat file taxon list. During puzzling the program 'reads' from the top down, carrying out pairwise comparisons between all possible combinations starting in reference to the top taxon.

- Rate heterogeneity: For uniform rates of sequence evolution (unlikely) $\gamma=0$, but higher, more probable γ values (maximum of 8) are also associated with longer sampling times. Given this trade-off, I used $\gamma=4$ setting for and kept analysis running times within realistic margins.
6. Confirm the above criteria to activate quartet puzzling.
 7. Name the resulting output file as appropriate. Quit TreePuzzle 5.0 program without saving: the newly named output file is preserved in the TreePuzzle folder.
 8. Open the accompanying file 'outfile' in a text application such as Word, SimpleText, or BBEdit. Note the Log likelihood (LnL) value present below the second cladogram diagram: more negative values represent increased statistical support.
 9. Open the named output file in TreeView, and handle as described with reference to TreeView files resulting from Bayesian analysis.
 10. Remove the infile, re-naming it with original name and .dat suffix for reference.

8.3.7 Comparative protein sequence analysis

I used PHYLIP (Phylogeny Inference Package) version 3.6 to calculate the degree of protein sequence divergence between compared taxa, relative to a defined reference sequence (Felsenstein 2001). Part of the output of this phylogeny inference program is a dot matrix alignment that graphically represents amino acid site conservation (dots) or substitution (single letter amino acid code) for all aligned sequences relative to the chosen reference.

PHYLIP program execution protocol

1. Save the relevant NEXUS alignment, including Ex/InSet criteria as .dat or PHYLIP 3.6 format. The reference sequence is the one at the top of the alignment.
2. Within the PHYLIP program, select Jones-Taylor-Thornton model parameters to calculate protein differences ('protein distances')
3. For computational execution of PHYLIP:
 - a) Open folder 'Phylip', and folder 'exe.2' within Phylip.
 - b) Rename the PHYLIP-format '.dat' file as 'infile'.
 - c) Transfer infile to the folder exe.2.
 - d) Open the program ProtDist, editing optional criteria to (i) replace the previous outfile and (ii) 'print out data (i.e. dot matrix) at the start'. Run program.
 - e) After running analysis, Quit and rename the 'outfile' output generated in the ProtDist folder.
 - f) Open re-named outfile in a text application such as Word or BBEdit, to view protein distance dot matrix alignment as in Figure 8.3.7a below.

g) Remove the infile from the ProtDist folder, and replace its original .dat filename.

```
Dm D11 .....KPRTIYSS LQLQQLNRRF QRTQYLALPE RAELAASLGL TQTQVKIWFQNRSSKYKKMM
Ag D11 .....
Jc D11 .....
Ba D11 .....
Tc D11 .....
Ci D1x .....A.....Q.....T.....C..L.
Tu D11 .....I.....L..K..N..Q
Pf D1x .....VL
Sk D11 .....C..L
Cs D11 .....L
Bf D11 .....T. F.....Q.....L.
Mm D1x1 .....A.....Q.....K..F..L.
Mm D1x6 .....A..H...Q.....K..F..LL
Dr D1x4 .....F..AA.Q...N.....K..L..I.
Mm D1x5 .....F..AA.Q...K.....K..I..I.
Mm D1x2 .....F..AA.Q...K.....F..W
Mm D1x7 .....H.DQ...H.....Q...Q.....K.....LL
```

Figure 8.3.7a Raw dot matrix output for Distal-less ProtDist analysis

8.4 *In situ* hybridization

8.4.1 DIG labeled ssRNA probe synthesis

I used purified plasmid DNA carrying inserts of putative *Dll*-regulatory genes as raw material for synthesis of sense (5' to 3') and anti-sense (3' to 5') single stranded RNA probes for subsequent *in situ* hybridisation in fixed, whole mount *Tetranymphus* embryos. Probes were labelled with digoxigenin using a mix of digoxigenin conjugated NTPs (Roche DIG RNA labelling mix). The general process involved:

- Linearization of plasmid, via a restriction enzyme whose cut site is only present in the multiple cloning site (MCS) at one end of the insert
- Phenol-chloroform extraction and ethanol precipitation of linearised DNA
- Transcription reaction, combining DIG-labelling and single-stranded RNA synthesis
- Precipitation of mRNA probe transcripts
- Re-suspension of mRNA probe transcripts

Protocol based on the Roche Kit protocol for mRNA probe synthesis

1. Ensure RNase free working conditions, using gloves, sterile filter tips, tubes etc. and cleaning all surfaces with mild detergent followed by 'RNase Zap', a solution designed to eliminate RNases.

2. ***Plasmid linearization*** Choose a restriction enzyme (RE) appropriate to cut within the MCS of the particular plasmid vector, as close as possible to the relevant end of the insert. The RE must be a non-cutter with respect to the insert sequence, and digest the plasmid DNA such that a 5' overhang is created to enable efficient attachment of the RNA polymerase during transcription. REs that satisfy these requirements for inserts in the plasmids relevant to this project, i.e. pGEM-T Easy and pBK-CMV, are tabulated below:

VECTOR PLASMID	INSERT ORIENTATION	RE FOR (+)RNA STRAND SYNTHESIS	RE FOR (-)RNA STRAND SYNTHESIS
pGEMTEasy	T7 5'→3' SP6	Spe1, Nde1	Nco1
pGEMTEasy	T7 3'→5' SP6	Nco1	Spe1, Nde1
pBK-CMV	T7 5'→3' T3	EcoR1, Spe1	Xho1, Not1
pBK-CMV	T7 3'→5' T3	Xho1, Not1	EcoR1, Spe1

Measure [plasmid DNA] with a spectrophotometer, diluting 1µl plasmid DNA in 100µl DEPC H₂O and correct readings according to the necessary factor. Absorption at a wavelength of 280nm (A₂₈₀) x50 = [DNA]µg/ml, and an A₂₈₀/A₂₆₀ ratio of 1.8 indicates optimal DNA purity. Between 1 and 3µg DNA is sought for each restriction digest, with reagents combined in a 1.5ml eppendorf (recipe below) and incubated at 37°C for at least 90'.

Restriction digest recipe: 10x buffer 1µl
 RE 1µl
 1-3µg plasmid DNA
DEPC H₂O
 10µl total

3. ***DNA extraction and precipitation*** Add 90µl ddH₂O to the 10µl restriction digest, for a final volume of 100µl. Add an equal volume (i.e. 100µl) of 1:1 phenol-chloroform, immediately inverting the tube at least 18 times to ensure thorough mixing. Spin tube and contents for 3' at maximum speed (13,000 rpm) in a micro-centrifuge chilled to 4°C. After spinning, carefully remove the upper aqueous layer from the biphasic mixture that forms, and transfer it to a new tube. Add an equivalent volume (100µl) of chloroform, inverting the tube and spinning again for 3' at 13,000 rpm and 4°C. Repeat transfer of the upper aqueous layer that separates out, into a new 1.5ml tube. Add 2x volume (i.e. 200µl) of absolute ethanol (99.96%) chilled to minus 20°C. Immediately invert the tube 5x then put at minus 20°C for 3' prior to spinning at 4°C for 25' at 13,000 rpm. Remove as much ethanol as possible, taking care not to disturb the DNA pellet. To wash the pellet, add 500µl 70% ethanol chilled to minus 20°C, inverting the tube 5x before a spin of 4' at 13,000rpm and 4°C. After this final spin, carefully remove as

much ethanol as possible and dry the pellet for 15-20', with a Speedy-Vac to provide centrifugation, vacuum and a held temperature of 50°C.

4. **ssRNA transcription** For transcription of single-stranded RNA, the following reagents are added to the dried pellet of precipitated DNA, for 20µl final volume:

Transcription recipe:

- 13µl DEPC H₂O
- 2µl 10x transcription buffer
- 2µl 10x DIG RNA labelling mix (10x)
- 2µl RNA polymerase* (SP6/T3/T7)
- 1µl RNase inhibitor
- 20µl total

*RNA polymerases appropriate to catalyse transcription from differently oriented inserts in the two vectors pGEMTEasy and pBK-CMV, are tabulated below:

VECTOR PLASMID	INSERT ORIENTATION	RNA POLYMERASE FOR (+) STRAND	RNA POLYMERASE FOR (-) STRAND
pGEMTEasy	T7 5'→3' SP6	T7	SP6
pGEMTEasy	T7 3'→5' SP6	SP6	T7
pBK-CMV	T7 5'→3' T3	T7	T3
pBK-CMV	T7 3'→5' T3	T3	T7

Briefly vortex the reactants to mix, then spin down all tube contents by micro-centrifugation for a few seconds. Incubate the transcription reaction at 37°C for at least 2hrs. Stop the reaction by addition of 2µl 0.2M pH8.0 EDTA.

5. **mRNA extraction and precipitation** To the 20µl transcription reaction add 0.1x (10%) volume of 4M LiCl (i.e. 2µl) and 2.5x volume of minus 20°C ethanol (i.e. 50µl). Mix well by pipetting gently up and down, then transfer the tube to minus 80°C for at least 30'. Spin for 15' at 13,000rpm and 4°C, then remove the supernatant ethanol-salt mixture without disturbing the pellet that will have formed. Wash the pellet carefully with 50µl 70% EtOH diluted in DEPC H₂O and pre-chilled to minus 20°C. Spin for 15' at 13,000rpm and 4°C, then remove as much ethanol as possible before drying the pellet for 15-20', rotating under vacuum at 50°C in SpeedyVac.
6. **mRNA probe resuspension and storage** Once dry, resuspend the ssRNA pellet in either 49µl DEPC H₂O or 49µl hybridisation buffer, with 1µl RNase inhibitor to protect the highly unstable RNA from enzymatic degradation. Aliquot and transfer to minus 20°C for storage, or to minus 80°C for more long-term (>12 month) storage.

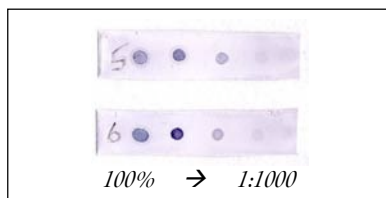
8.4.2 RNA hybridization to test probe labeling efficiency

After synthesis of ssRNA probes, I confirmed the efficacy of the digoxigenin (DIG) labelling reaction by binding serially diluted dots of probe samples to a nitrocellulose membrane, hybridising the 'dot blots' with digoxigenin-alkaline phosphatase (DIG-AP)-conjugated antibody and observing the rate and intensity of the subsequent staining reaction. Protocol and reagent details are given as follows:

RNA probe dot blot protocol

1. *Preparing dot blots* Cut nitrocellulose membrane (Hybond-NKX optimised for NA transfer) into 6x20mm strips. Label at one end as appropriate, with pencil. Make serial dilutions of each probe to be tested, in as small a volume of DEPC H₂O as possible, dilutions ranging from pure probe to 1:1000. Apply 0.75µl dots of each dilution in order, onto the nitrocellulose strips. Allow strips to dry, then auto-crosslink to bind the RNA samples irreversibly to the membrane. Using forceps, transfer the prepared strips into 2ml eppendorf tubes for hybridisation.
2. *Hybridisation with DIG-AP* Add 1.5ml Buffer 1 for 1' to hydrate membranes. Replace with Buffer 2 and incubate in a 42°C water-bath for 20'. Replace with 400µl Roche anti-DIG, alkaline phosphatase-conjugated F_{ab} antibody fragments, diluted at 1:2000 in Buffer 1, incubating with rocking at RT for 1hr.
3. *Washing* Transfer the strips of membrane to one 50ml Falcon tube and wash 3x 10' with 40ml Buffer 1, incubating at room temperature on rollers. Wash 3x 10' with 40ml Buffer 3, then return the strips to 2ml eppendorf tubes.
4. *Signal detection* To each tube add 100µl staining solution, with 3.6µl NBT-BCIP mix diluted in 100µl Buffer 3. (Chromagens nitro-blue-tetrazolium chloride (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) are pre-combined in Roche NBT-BCIP stock solution at 18.75mg/ml and 9.4mg/ml respectively.) Incubate tubes at room temperature with rocking, ensuring that strips are being completely covered by the staining solution.

5. Observe development of staining, allowing the reaction to proceed until a clear signal is visible, or for a maximum of 15'. The sample shown below exemplifies the form of staining indicative of a positive result, with stain intensity clearly decreasing with increased probe dilution factor:



6. *Terminating the reaction* To stop the colour reaction, replace staining solution with 1.5ml termination buffer and invert the tubes several times. Pour off the termination buffer and allow membranes to dry before storage at 4°C, protected from moisture and light.

Reagents for dot blot protocol

Buffer 1 (10x): 121.0g Tris pH8.5 (1M final conc.)
 58.4g NaCl (1M final conc.)
 4.07g MgCl₂ (0.02M final conc.)
 ddH₂O 1L total volume
 → autoclave; allow to cool
 5ml Triton-X100 (0.5% final vol.)
 → store at 4°C up to 3 months

Buffer 2 (10x): 121.0g Tris pH8.5 (1M final conc.)
 58.4g NaCl (1M final conc.)
 4.07g MgCl₂ (0.02M final conc.)
 ddH₂O 1L total volume
 → autoclave; allow to cool
 5ml Triton-X100 (0.5% final vol.)
 2g BSA (0.2% final w/vol.)
 → store at 4°C up to 3 months

Buffer 3 (staining buffer): 100ml 1M Tris HCl pH9.5 (0.1M final conc.)
 5.84g NaCl (0.1M final conc.)
 10.16g MgCl₂ (0.05M final conc.)
 ddH₂O 1L total volume

Termination buffer:	1ml 1M Tris pH8.5	(final conc. 20mM)
	0.5ml 0.5M EDTA	(final conc. 5mM)
	48.5ml ddH ₂ O	50ml total volume

8.4.3 Whole-mount mRNA *in situ* hybridization protocol

The protocol for *in situ* hybridization on *Tetranychus* embryos was adapted primarily from that in Patel (1994) for whole-mount *Drosophila* antibody staining, also taking into account techniques described for *Tetranychus* in Dearden, Donly and Grbic (2002), for the mite *Archegozetes longesitosus* in Telford and Thomas (1998) and for the spider *Cupiennius salei* in Damen and Tautz (1998). Protocol and reagents are as follows:

Protocol for spider mite in situ hybridisation

1. Transfer fixed embryos, in 100µl volume of hybridization buffer ('hyb'), to a 20mm diameter Pyrex staining well. To protect from particulates and light, keep the staining well plate in an aluminium-foil-wrapped Petri dish with lid, except when changing solutions.
2. Pre-hybridise by incubation at 58°C for 2hrs. Add RNA probe to the 100µl hyb + embryos, at approximately 0.1-1ng/µl. (Experimentation with multiple ribo-probe concentrations may be necessary to achieve optimal signal:background staining; I found that 1µl mRNA diluted 10x per 100µl *in situ* reaction often worked well.) Incubate the hybridisation reaction(s) at 58°C for 12-16hrs.
3. To rinse, carefully remove hyb buffer with G30.5 syringe needle and replace with wash buffer. Similarly using the fine syringe to remove solutions without disturbing the embryos, wash at 58°C, performing 5x 15'-20' washes. Rinse with PT at room temperature, then wash for 30' with PT at room temperature on a gently rotating shaker.
4. Replace PT with secondary antibody - Roche's anti-DIG, alkaline phosphatase-conjugated (anti-DIG-AP) F_{ab} antibody fragments - diluted at 1:1000 in PAT, and incubate overnight at 4°C on a gently rotating shaker.
5. Rinse with PT, then wash 5x in PT, 15' per wash, at room temperature on gently rotating shaker. Rinse 2x in AP buffer, then wash in AP buffer for 10' at room temperature.
6. Replace AP buffer with 100µl NBT-BCIP mix diluted in AP buffer. Roche's NBT-BCIP stock solution (NBT 18.75mg/ml + BCIP 9.4mg/ml) was used at 3.6µl per 100µl AP buffer, for final concentrations of 0.9µl NBT @ 75mg/ml and 0.7µl BCIP @ 50mg/ml. Incubate with gentle rocking at room temperature, observing development of the staining reaction until the

colour reaction reaches an optimum signal intensity (relative to any background ‘noise’ signal).

7. Stop the colour reaction by rinsing 3x in PT, then washing 2x 10’ in PT. To reduce background staining it may be fruitful at this point to dehydrate the embryos through serial dilutions of methanol in PT, incubating in 100% methanol until a more desirable signal:noise balance is achieved. If the dehydration step is carried out, embryos are rehydrated to 100% PT prior to DAPI nuclear counterstaining (c.f. section 8.4.4) or transfer to glycerol.

Reagent formulae

REAGENT	COMPOSITION	E.G. RECIPE
Hybridisation buffer	50% formamide 4x SSC 1x Denhardt’s solution 250µg/ml yeast RNA ²⁶ 50µg/ml heparin 0.1% Tween-20 5% dextran sulphate ddH ₂ O	25ml 10 ml 20x 1ml 50x 3.125ml 20mg/ml 125µl 20mg/ml 50µl 5ml 50% <u>5.7ml</u> 50ml total
Wash buffer	50% formamide 2x SSC 0.1% Tween-20 ddH ₂ O	25 ml 5ml 20x 50µl <u>20ml</u> 50ml total
PT	1x PBS 0.1% Triton	
PBT	1x PT 1% BSA (bovine serum albumin)	50ml PT 500µg BSA
AP buffer	0.1M (100mM) TRIS pH9.5 0.1M (100mM) NaCl 0.05M (50mM) MgCl ₂ 0.1% Tween-20 solvent: ddH ₂ O	1ml 1M TRIS pH9.5 1ml 1M NaCl 500µl MgCl ₂ 10µl Tween-20 8.5ml ddH ₂ O
Denhardt’s	1% Ficoll <i>Type 400, Pharmacia</i> 1% polyvinyl-pyrrolidone 1% BSA (bovine serum albumin) ddH ₂ O	1g 1g 1g <u>97ml</u> 100ml total
SSC 20x	NaCl Sodium citrate ddH ₂ O + NaOH to pH7 ddH ₂ O to final volume	17.5g 88.2g 800ml to 1L total

²⁶ Alternative blocking agents: 250µg/ml tRNA + 250µg/ml boiled ssDNA

8.4.4 DAPI nuclear counter-staining

About 4'-6-diamidino-2-phenylindole

DAPI (4'-6-diamidino-2-phenylindole, $C_{16}H_{15}N_5$) is a cell-permeable molecule that forms a stable UV-fluorescent complex with double stranded DNA, binding preferentially to AT base-pair acceptor sites and hydrogen bonding within the nucleic acid groove (Figure 8.4.4a) (Green 1990; Sigma 1997). The fluorescence of DAPI when in complex with DNA is enhanced by a factor of 20, making this fluorophore an excellent marker for nuclei in diverse cytochemical studies. It is typically used in its di-hydrochloride form ($C_{16}H_{15}N_5 \cdot 2HCl$), permeabilised tissues being exposed to concentrations in the region of 10 $\mu g/ml$ (Green 1990).

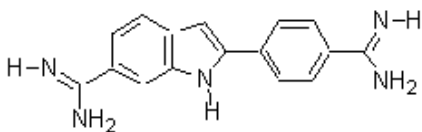


Figure 8.4.4a Structural formula of 4'-6-diamidino-2-phenylindole (DAPI)

The morphology and distribution of DAPI-stained nuclei are clearly observable by fluorescence microscopy at an excitatory wavelength of 350nm: for absorption and emission spectra see Figure 8.4.4b (Du et al. 1998).

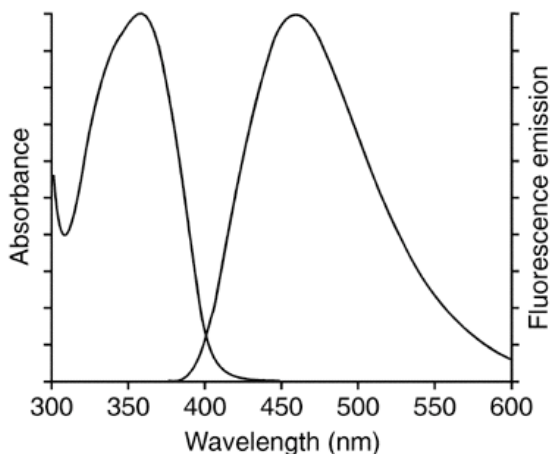


Figure 8.4.4b: Absorption and fluorescence emission spectra of DAPI bound to DNA

The fluorescent marker DAPI acts as a counter-stain to dark *in situ* mRNA or antibody staining patterns, revealing both early zygotic cleavage patterns and later embryo morphology. I used DAPI

staining as standard after *in situ* and antibody detection in whole mount spider mite embryos, according to the protocol described in the following paragraph.

Spider-mite DAPI staining protocol

I exposed embryos to DAPI (Sigma 4',6-diamidino-2-phenylindole dihydrochloride hydrate) either immediately after terminating the chromagen-phosphatase reaction with PT or after rehydrating embryos from methanol (Sigma 1997). I first diluted the 1mg/ml Sigma DAPI stock solution in PT to a working concentration of 100µg/ml, then added 1µl DAPI working solution (equivalent to 0.1µg DAPI) to 100µl PT + embryos in a staining vessel, for a final concentration of 1µg/ml. The staining dish was incubated for 15'-30' in a foil-covered, light-excluding Petri dish, with gentle shaking at room temperature. After dye absorption, I rinsed embryos with PT, then washed 2-3x for 10' in PT (see section 8.4.3 for reagent mixes). DAPI stained embryos were subsequently transferred to glycerol for photo-microscopy, or dehydrated into 100% methanol for storage at minus 20°C.

8.5 Immuno-histochemical staining

8.5.1 Antibody staining in *Tetranychus urticae*

Tetranychus embryos were antibody stained after fixation under the conditions described in section 8.1.4, which are milder than for embryos fixed for *in situ* (section 8.1.3). The final protocol adapted from that developed by Telford (1998) for *Archegozetes longisetosus* and is based on methods originally described in Patel (1994) (Patel 1994a; Patel 1994b; Telford and Thomas 1998a). I used both HRP- (horse radish peroxidase) and AP- (alkaline phosphatase) conjugated secondary antibodies, with respective DAB (di-aminobenzoate) and NBT-BCIP chromogen reactions. I adapted distinct protocols to suit each enzyme conjugate-chromogen pair (HRP-DAB vs. AP-NBT/BCIP), the different steps and reagents distinguished within protocols I and II as follows:

Protocol I: peroxidase-DAB antibody staining

- Handle fixed embryos in a Pyrex staining well plate inside a foil covered Petri dish. A G30.5 syringe is used to change solutions aided by a dissecting microscope, under which staining reactions are also observed.
- Block embryos for 30' in 200µl PBT followed by 30' in 200µl PBT-NGS, at room temperature on a gently rotating shaker. After blocking, add a minimal volume of primary antibody, diluted in PBT-NGS at multiple test concentrations or at a specific pre-ascertained concentration (see 8.5.4 for values pertinent to each antibody tested in this project). Incubate embryos overnight at 4°C with gentle rotation. *Negative control* embryos (n=~10) are set aside and incubated in only PBT-NGS at this stage, as a test for any endogenous staining or other response to the reagents, independent of antibody binding to target cellular proteins. *Positive control* embryos (n=~10) are in most cases incubated in a rabbit antibody against phospho-Histone-III at 1:400 in PBT-NGS. α -p-Histone-III is a reliable and robust marker for mitotic cells in many metazoans, being bound by elements of dividing centromeres, and hence is an ideal positive control antibody.
- Remove primary antibody by rinsing 3x with PT and washing 5x for 20' with PT, at room temperature and with gentle rotation. Block embryos for 1hr in PBT-NGS at room temperature. Add 100µl secondary antibody in PBT-NGS, the type and concentration of enzyme conjugate corresponding to whether the primary antibody is raised against mouse or rabbit antigens; anti-rabbit peroxidase is used at 1:500, and anti-mouse peroxidase at 1:400. Incubate overnight as before at 4°C.
- Remove secondary antibody and excess protein serum by rinsing 3x with PT and washing 2x 20', followed by 2x 10' washes with PTw, lowering [detergent] in preparation for the DAB reaction.
- Defrost an aliquot of DAB from minus 20°C storage, and add 200µl to each staining well before addition of 10µl hydrogen peroxide/H₂O₂ solution (3µl H₂O₂: 1ml PTw). Allow reactions to proceed in darkness (in foil-coated Petri) and on a gently rotating shaker, inspecting embryos every few minutes under a dissection microscope to track reaction progress as brown chromogenic substrate precipitates.
- Once an adequate signal is achieved, with minimal interference from background staining, terminate the reaction by rinsing 4x and washing 2x for 10' with PTw. Embryos are then either (i) dehydrated into methanol for storage at minus 20°C, or (ii) nuclear stained with DAPI (c.f. section 8.4.4), prior to transfer into glycerol for photomicroscopy (section 8.6).

Protocol II: AP-NBT/BCIP antibody staining

- In a Pyrex staining well and foil-covered Petri dish as before, block embryos for 30' in 200µl PBT followed by 30' in 200µl PBT-NGS, at room temperature on a gently rotating shaker. After blocking, add a minimal volume of primary antibody, diluted in PBT-NGS to multiple test concentrations or to a predetermined optimum (see section 8.5.4 for specific values). Incubate experimental and control embryos in primary antibody overnight at 4°C with gentle rotation. Set aside *negative control* embryos (n=~10) at this stage, incubated in only PBT-NGS as a test for endogenous cellular responses to any of the reagents, independent of antibody binding to target proteins. *Positive control* embryos (n=~10) are in most cases incubated in a rabbit antibody against phospho-Histone-III (pHisIII) at 1:400 in PBT-NGS, as in protocol I.
- Remove primary antibody by rinsing 3x with PT and washing 5x for 20' with PT, at room temperature and with gentle rotation. Block embryos for 1hr in PBT-NGS. Add secondary antibody, the type and concentration of enzyme conjugate corresponding to primary antibody raised against either mouse or rabbit antigens. PBT-NGS is diluted to 1:400 for anti-rabbit-alkaline phosphatase and 1:300 for anti-mouse-alkaline phosphatase. Incubate embryos in secondary antibody overnight at 4°C, with gentle rotation as before.
- Remove secondary antibody and excess protein serum by rinsing 3x and washing 5x 20' with PT.
- Rinse 2x and wash for 10' in AP buffer at room temperature. Replace with 100µl NBT-BCIP-AP buffer. Use Roche NBT-BCIP stock solution (NBT 18.75mg/ml + BCIP 9.4mg/ml) at 3.6µl per 100µl AP buffer. Incubate with gentle rocking and protection from light at room temperature, observing precipitation of stain at intervals under a dissecting microscope until the colour reaction gives an optimum signal:noise intensity. Depending on the probe, optimum signal may occur after ~10 minutes or after several hours. (When background signal is a problem, lower [probe] and longer stain exposure at lower temperatures (e.g. 4°C) often proved helpful.)
- Stop the phosphatase reaction by rinsing 4x and washing 3x 10' in PT. To further reduce background staining it is sometimes beneficial to dehydrate embryos through serial dilutions of methanol, incubating in a final concentration of 100% methanol until a better signal:noise balance is achieved. Embryos must be rehydrated into 100% PT prior to further counterstaining with DAPI or transfer to glycerol.

Reagents

REAGENT/SOLUTION	COMPOSITION	WORKING VOLUMES
PT	1x PBS 0.1% Triton	N/A
PT _w	1x PBS 0.2% Tween-20	N/A
PBT	1x PT 1% BSA (bovine serum albumin)	50ml PT 500µg BSA
PBT-NGS	1x PBT 5% NGS (normal goat serum)	950µl PBT 50µl NGS
DAB	1x	200µl
H ₂ O ₂ staining mix	3µl H ₂ O ₂ 1ml PT _w	10µl
AP buffer	0.1M (100mM) TRIS pH9.5 0.1M (100mM) NaCl 0.05M (50mM) MgCl ₂ 0.1% Tween-20 solvent: ddH ₂ O	1ml 1M TRIS pH9.5 1ml 1M NaCl 500µl MgCl ₂ 10µl Tween-20 8.5ml ddH ₂ O
NBT-BCIP-AP buffer	0.68µg/µl NBT 0.34µg/µl BCIP 1x AP buffer	100µl AP buffer 3.6µl NBT(18.75mg/ml)+BCIP (9.4mg/ml) stock mix

8.5.2 Antibody staining in *Cupiennius salei*

To complement work on *Tetranychus-opp*, embryos of the spider *Cupiennius salei* were stained with anti-(rabbit)-phospho-Smad1, conjugated to anti-(rabbit)-alkaline phosphatase. Spider embryos at various stages of development were obtained from W.G.M. Damen and A. Stollewerk (Universität zu Köln), fixed and in methanol.

The protocol for staining was essentially the same as that described for *Tetranychus*, but the much larger egg diameter of *Cupiennius* (1200µm vs. ~100µm for *Tetranychus*) necessitated modified means of handling embryos and solution volumes. In addition, prior to commencing the staining protocol proper, embryos required rehydration and manual removal of vitelline membranes. Details pertinent to these modifications are given below:

- Embryos are moved between tubes and dissection/staining vessels using a 1ml Gilson pipette, the plastic pipette tip trimmed to widen the aperture to ~1.5mm for uptake of spider embryos without causing shear tissue damage.
- Fixed embryos in methanol are transferred to a black watch-glass (better for visibility than colourless glass) and the vitelline membranes, that persist throughout fixation, are manually dissected off under a dissecting light microscope, using a 30.5G syringe needle and a pair of fine forceps. (The physical properties of the vitelline membrane are more amenable to

dissection in methanol than in PT, although regularly replacing evaporated alcohol and keeping a glass lid on the watch glass as much as possible, is required.)

- Rehydration from methanol is carried out with embryos in 1.5ml eppendorf tubes, incubated for 5' with gentle rotation at room temperature, in 1ml of serial dilutions from pure MeOH to pure PT (i.e. 3:1 MeOH:PT, 1:1 MeOH:PT, 1:3 MeOH:PT, 100% PT).
- De-vitellinised and rehydrated embryos are transferred to the chamber(s) of an aluminium foil-covered 24-well plate, one experimental batch per chamber. During staining, solutions are added using a 1000 or 200µl Gilson pipette, and removed using a G19 syringe needle so as not to disturb the embryos yet remove as much liquid as possible.
- Antibody staining with anti-pSmad1/anti-rabbit-AP/NBT-BCIP is carried out as described for *Tetranychus*, except that larger volumes and longer washes were needed for the *Cupiennius* protocol. Solution volumes of 250-400µl are adequate to cover embryos in 1 chamber of the 24-well plate, 4x 30' PT washes are performed prior to adding secondary antibody, and 5x 20' PT washes prior to developing staining patterns with NBT-BCIP.
- Subsequent DAPI staining and/or transfer to 95% glycerol is carried out exactly as for *Tetranychus*, merely substituting larger volumes where and when necessary.

8.5.3 Primary and Secondary antibodies

Salient details for primary and secondary antibodies used during attempted spider mite protein detection are tabulated in sections (i) and (ii) below:

i) Primary antibodies

ANTIBODY TARGET	SOURCE	ANTIGEN TYPE	MONOCLONAL / POLYCLONAL	TYPICAL WORKING CONC.
phospho-Histone-III	Sigma	Rabbit	Mono	1:400
Distal-less	G. Panganiban, University of Wisconsin	Rabbit	Poly	1:200
dual-phosphorylated MAPK (dpERK)	Sigma (Gabay, L. 1997)	Mouse	Mono	1:200
phospho-Smad1	C.-H. Heldin, Ludwig ICR, Uppsala	Rabbit	Poly	1:2000

UbdA (FP6.87)	M. Averof, ICMB Crete	Rabbit	Poly	1:10
Wingless	Developmental Studies Hybridoma Bank, Iowa	Mouse	Mono	1:50
Armadillo (N2)	DSHB	Mouse	Mono	1:100
Engrailed (4D9)	Patel, N.	Mouse	Mono	1:50

ii) Secondary antibodies

ANTIBODY TYPE	ENZYME CONJUGATE	TYPICAL WORKING CONCENTRATION
Rabbit	Alkaline Phosphatase (AP)	1:400
Mouse	AP	1:300
Rabbit	peroxidase	1:500
Mouse	peroxidase	1:300

8.6 Microscopy and Imaging

8.6.1 Mounting stained embryos

Transfer to 95% glycerol

Before whole-mounting *Tetranynchus* or *Cupiennius* embryos on slides, the embryos are gradually transferred to a 95% glycerol ($C_3H_8O_3$) solution. The viscosity of a 95% glycerol medium is ideal for mounting embryos such that they can be rotated to a desired position when force is applied, and yet remain fairly static when external forces are removed.

Once mRNA, antibody or DAPI staining was complete, PT/PTw was replaced with 50% glycerol (at 1:1 ratio with PT) and incubated for at least 1hr at room temperature with gentle rocking. This step was repeated with a 75% glycerol solution, this being finally replaced by 95%. On addition of 95% glycerol, embryos were incubated overnight at 4°C with gentle rocking, to ensure adequate embryo equilibration to the dense solution.

Mounting Tetranychus embryos for microscopy

Using a 10µl pipette tip, *Tetranychus* embryos were transferred individually from 95% glycerol onto a 72x26mm glass slide (Gold Star, Chance Propper Ltd). A few extra drops of glycerol was added, and the embryo centred with the aid of a dissecting microscope and G30.5 syringe needle. Very small pieces of plasticine were attached to form 'feet' at the four corners of a square 22x22mm borosilicate glass coverslip (BDH Cover Glass, thickness no. 1). The coverslip was lowered carefully onto the glycerol drop + embryo, using fine forceps to lay it from one side and taking care to avoid forming air-bubbles.

Preparing Cupiennius embryos for microscopy

Cupiennius embryos in 95% glycerol were amenable to photomicroscopy after transferring embryo(s) of interest from their staining well directly into a small colourless watch glass, supported by a further 500µl-1ml glycerol. Individual embryos were selected by inspection with a dissecting microscope, transferred by means of a cut-off 1ml pipette tip, and gently rotated to the viewing angle (or angles) desirable for subsequent image capture using a syringe needle.

8.6.2 Photomicrography

i) Photomicrography of fixed embryos

Tetranychus

Nomarski/DIC and UV fluorescence images were captured on a Zeiss Axiophot microscope linked to a Leica DFC 300F digital camera, and processed using Leica FireCam 1.3.0 or Openlab 3.1.5 software for Macintosh. Multidimensional images (i.e. Z-plane stacks) through *Tetranychus* embryos were acquired using semi-automatic motorised microscope (Zeiss Axioskop 2 mot plus) linked to a Zeiss AxioCam MRm digital camera and processed *via* AxioVision software for (evil) PC. When DAPI fluorescence was being visualized, the UV halogen bulb was allowed to warm up for a few minutes prior to use, and – for safety purposes due to the volatility of the bulb contents - to cool down for at least 30' after switching off before turning the bulb on again. As mentioned in section 8.4.4, DAPI stained embryos are best viewed through narrow band fluorescence filters that excite the fluorophore close to its optimal absorption wavelength of 350nm (Green 1990).

Cupiennius

DIC or transmitted light images of *Cupiennius* embryos, being so much larger than mite embryos, were visualized directly under a dissection microscope (Leica MZ FLIII), attached to a Leica DFC 300FX digital camera and Leica FireCam 1.3.0 software for image processing.

ii) Photomicrography for live spider mite embryos and hatchlings

Individual spider mite embryos selected at the 1 or 2-cell-stage were mounted in a drop of 1x PBS, and the cover slip sealed with a breathable silicon oil to prevent dessication and yet allow sufficient respiration. These earliest cleavage embryos were laterally illuminated on a Zeiss Axioscop, and time-lapses recorded with OpenLab software. Time-lapses were allowed to continue for 12-24 hours, effectively recording from a single angle the pattern of zygotic cleavage, blastula formation and phases of germdisc and early germband morphogenesis. Live eggs, hatchlings and adults were observed on the underside of a bean leaf using a Zeiss Stemi ZV11 dissection microscope. Still images, including a succession depicting hexapod hatchling emergence, were taken using the dissecting microscope linked to a Leica DFC 300F digital camera and Leica FireCam software.

8.6.3 Scanning electron microscopy

I made separate collections of embryos at six distinct developmental stages (listed below), followed by boiling, manual membrane dissection and formaldehyde fixation according to the protocol described in section 8.1.3. After fixation, I dehydrated embryos into pure ethanol and stored them at minus 20°C, awaiting preparation for SEM.

Developmental stages defined for investigation

- i) early cleavage – blastoderm
- ii) germ disc or pre-disc asymmetry in blastoderm cell density
- iii) early limb buds
- iv) mid limb development
- v) late (jointed) limb development
- vi) hexapod hatchlings and 8-legged nymphs

Preparing embryos for SEM

I transferred fixed embryos, with ethanol, to a 100µm-mesh-based basket. Baskets containing embryos at stages (i) to (vi) were loaded into a canister and placed into a critical point drying chamber. A balance in air pressure between internal and external tissue environments of the samples is slowly reached, then at this ‘critical point’ pressure is suddenly increased so that CO₂ within the

canister sublimates instantaneously to CO₂ gas, completely removing all moisture from (and conferring a degree of brittleness to) the tissue samples.

Dried embryo specimens from each developmental stage were transferred to adhesive carbon discs and attached to metal buttons that fit inside the SEM chamber. Carbon-disc mounted samples were integrated into the anode (+) of a vacuum chamber with a gold-coated cathode (-) on the opposite surface. A low vacuum (10⁻³ torr) is created, into which Argon gas is released. When a high voltage is passed through the system, Ar molecules become ionized and these positively charged particles bombard the Au cathode surface, displacing target Au atoms (cations) and electrons (-) in all directions. Displaced particles migrate in a beam of plasma to the oppositely charged anode surface, and coat the surface of the specimens fixed at the anode. This process of 'sputter-coating' was used to coat the dried *Tetranychus* embryos in a thin film of gold (Au), in preparation for exposure to the accelerated electron beam of the scanning electron microscope.

SEM microscopy

Scanning electron microscopy was carried out on the Department of Anatomy scanning electron microscope, images captured using software for PC and saved in TIF format for further editing or processing.

8.6.4 Image processing

I processed photo-microscopy images captured with OpenLab or Leica software, using Adobe Photoshop version 7.0 to rotate embryo orientations and improve contrast etc. where appropriate. I used Adobe Illustrator version 10 for Macintosh to further annotate or incorporate processed images within figures.

REFERENCES



Lycose Narbonnaise

- Abzhanov, A., and T. C. Kaufman. 1999. Homeotic genes and the arthropod head: Expression patterns of the *labial*, *proboscipedia*, and *Deformed* genes in crustaceans and insects. PNAS USA 96:10224-10229.
- Abzhanov, A., and T. C. Kaufman. 2000a. Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. Dev 127:2239-2249.
- Abzhanov, A., and T. C. Kaufman. 2000b. Evolution of distinct expression patterns for *engrailed* paralogues in higher crustaceans (Malacostraca). Dev Genes Evol 210:493-506.
- Abzhanov, A., and T. C. Kaufman. 2000c. Homologs of Drosophila appendage genes in the patterning of arthropod limbs. Dev Biol 227:673-689.
- Abzhanov, A., A. Popadic, and T. C. Kaufman. 1999. Chelicerate Hox genes and the homology of arthropod segments. Evol & Dev 1:77-89.
- Adoutte, A., G. Balavoine, N. Lartillot, and R. de Rosa. 1999. Animal Evolution: the end of intermediate taxa? TIG 15:104-108.
- Adoutte, A., G. Balavoine, N. Lartillot, O. Lespinet, B. Prud'homme, and R. de Rosa. 2000. The new animal phylogeny: reliability and implications. PNAS USA 97:4453-4456.
- Ahlquist, P. 2002. RNA-dependent RNA polymerases, viruses, and RNA silencing. Science 296:1270-1273.
- Akam, M. 1987. The molecular basis for metameric pattern in the Drosophila embryo. Development 101:1-22.
- Akam, M. 1998. Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. Int J Dev Biol 42:445-451.
- Akam, M. 2000. Arthropods: Developmental diversity within a (super)phylum. PNAS 97:4438-4441.
- Akam, M., and A. Martinez Arias. 1985. The distribution of *Ultrabithorax* transcripts in *Drosophila* embryos. EMBO J 4:1689-1700.
- Akiyama-Oda, Y., and H. Oda. 2003. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. Development 130:1735-1747.
- Akiyama-Oda, Y., and I. Oda. 2006. Axis specification in the spider embryo: *dpp* is required for radial-to-axial symmetry transformation and *sox* for ventral patterning. Development 133:2347-2357.
- Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. 1994. Molecular Biology of The Cell. Garland Publishing
- Alexandre, C., and J.-P. Vincent. 2003. Requirements for transcriptional repression and activation by Engrailed in Drosophila embryos. Dev 130:729-739.
- Aljamali, M. N., A. D. Bior, J. R. Sauer, and R. C. Essenberg. 2003. RNA interference in ticks: a study using histamine binding protein dsRNA in the female tick *Amblyomma americanum*. Insect Mol Biol 12:299-305.
- Alonso, C. R., and A. S. Wilkins. 2005. The molecular elements that underlie developmental evolution. Nature Reviews Genetics 6:709-715.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic Local Alignment Search Tool. J Mol Biol 215:403-410.
- Amit, S., A. Hatzubai, Y. Birman, J. S. Andersen, E. Ben-Shushan, M. Mann, Y. Ben-Neriah, and I. Alkalay. 2002. Axin-mediated CKI phosphorylation of beta-catenin at Ser 45: a molecular switch for the Wnt pathway. Genes & Dev 16:1066-1076.
- Anderson, D. T. 1973. Chapter 9: Chelicerates. Pp. 365-451. Embryology and phylogeny in annelids and arthropods. Pergamon Press.
- Anderson, D. T. 1999. Subphylum Chelicerata in D. T. Anderson, ed. Invertebrate Zoology. OUP.
- André, H.M. and Ducarme, X. 2003. Rediscovery of the genus *Pseudotyrdeus* (Acari: Tydeoidea), with description of the adult using digital imaging. Insect Syst Evol 4: 373-379.
- Angelini, D. R., and T. C. Kaufman. 2004. Functional analyses in the hemipteran *Oncopeltus fasciatus* reveal conserved and derived aspects of appendage patterning in insects. Dev Biol 271:306-321.
- Angelini, D. R., and T. C. Kaufman. 2005a. Functional analyses in the milkweed bug *Oncopeltus fasciatus* (Hemiptera) support a role for Wnt signaling in body segmentation but not appendage development. Dev Biol 283:409-423.
- Angelini, D. R., and T. C. Kaufman. 2005b. Insect appendages and comparative ontogenetics. Dev Biol 286:57-77.
- Angelini, D. R., P. Z. Liu, C. L. Hughes, and T. C. Kaufman. 2005. Hox gene function and interaction in the milkweed bug *Oncopeltus fasciatus* (Hemiptera). Dev Biol 287 (2): 440-455.
- Apesteguía, S. and Zaher, H. 2006. A Cretaceous terrestrial snake with robust hindlimbs and a sacrum.

- Nature 440: 1037-1040.
- Appel, B., and S. Sakonju. 1993. Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in *Drosophila* embryos. EMBO J 12:1099-1109.
- Arora, K., M. Levine, and M. B. O'Connor. 1994. The *screw* gene encodes a ubiquitously expressed member of the TGF-beta family required for specification of dorsal cell fates in the *Drosophila* embryo. Genes & Dev 8:2588-2601.
- Arthur, W., T. Jowett, and A. Panchen. 1999. Segments, limbs, homology and co-option. Evol & Dev 1:74-76.
- Asano, M., Y. Emoris, K. Saigo, and K. Shiokawa. 1992. Isolation and characterisation of a *Xenopus* cDNA which encodes a homeodomain highly homologous to *Drosophila Distal-less*. J Biol Chem 267:5044-5047.
- Athanikar, J. N., H. B. Sanchez, and T. F. Osborne. 1997. Promoter selective transcriptional synergy mediated by sterol regulatory element binding protein and Sp1: a critical role for the Btd domain of Sp1. Molecular and Cellular Biology 17:5193-5200.
- Averof, M. 1994. HOM/Hox genes of a crustacean; Evolutionary implications. Department of Genetics. University of Cambridge.
- Averof, M. 2002. Arthropod Hox genes: insights on the evolutionary forces that shape gene functions. Curr. Opin. Genes & Dev. 12:386-392.
- Averof, M., and M. Akam. 1993. HOM/Hox genes of *Artemia*: implications for the origin of insect and crustacean body plans. Curr. Biol. 3:73-78.
- Averof, M., and M. Akam. 1995. Hox genes & the diversification of insect & crustacean body plans. Nature 376:420-423.
- Averof, M., and N. H. Patel. 1997. Crustacean appendage evolution associated with changes in Hox gene expression. Nature 388:682-686.
- Bajpal, R., K. Makhijani, P. R. Rao, and L. S. Shashidhara. 2004. *Drosophila* Twins regulates Armadillo levels in response to Wg/Wnt signal. Development 131:1007-1016.
- Baldauf, S. L. 2003. Phylogeny for the faint of heart: a tutorial. TIG 19:345-351.
- Barnes, Calow, Olive, Golding, and Spicer. 2001. Phylum Chelicerata in Blackwell, ed. The Invertebrates: A Synthesis.
- Barr, P. J. 1991. Mammalian subtilisins: the long-sought dibasic processing endoproteases. Cell 66:1-3.
- Barth, F. G., and J. Stagl. 1976. The slit sense organs of arachnids. Zoomorphologie 86:1-23.
- Bastianello, A. and Minelli, A. 2001. *Engrailed* sequences from four centipede orders: strong sequence conservation, duplications and phylogeny. Dev Genes Evol. 211: 620-623.
- Baulcombe, D. 2002. RNA silencing. Curr Biol 12:R82-R84.
- Beerman, A., D. G. Jay, R. W. Beeman, M. Hülkamp, D. Tautz, and G. Jürgens. 2001. The *Short antennae* gene of *Tribolium* is required for limb development and encodes the orthologue of the *Drosophila Distal-less* protein. Development 128:287-297.
- Beermann, A., M. Aranda, and R. Schröder. 2004. The *Sp8* zinc-finger transcription factor is involved in allometric growth of the limbs in the beetle *Tribolium castaneum*. Development 131:733-742.
- Bell, S. M., C. M. Schreiner, R. R. Waclaw, K. Campbell, S. S. Potter, and W. J. Scott. 2003. *Sp8* is crucial for limb outgrowth and neuropore closure. PNAS 100:12195-12200.
- Bely, A. E., and G. A. Wray. 2001. Evolution of regeneration and fission in annelids: insights from *engrailed* and *orthodenticle*-class gene expression. Development 128:2781-2791.
- Benfey, P. N. 2003. MicroRNA is here to stay. Nature 425:244-245.
- Bennet, R. L., S. J. Brown, and R. E. Denell. 1999. Molecular and genetic analysis of the *Tribolium Ultrabithorax* ortholog, *Ultrathorax*. Dev Genes Evol 209:608-619.
- Bielza, P. 2005. La plaga *Tetranychus urticae*. Pp. SEM study of *Tetranychus urticae* females and males. Sociedad Española de Entomología Aplicada: Divulgación, Madrid.
- Bienz, M. 2003. APC Quick guide. Curr Biol 13:R215-216.
- Blaxter, M. 2001. Sum of the arthropod parts. Nature 413:121-122.
- Blin, M., N. Rabet, J. S. Deutsch, and E. Mouchel-Vielh. 2003. Possible implication of Hox genes *Abdominal-B* and *abdominal-A* in the specification of genital and abdominal segments in cirripedes. Dev Genes Evol 213:90-96.
- Boudreaux, H. B. 1963. Biological aspects of some phytophagous mites. Ann Rev Entomol 8:137-154.
- Boutla, A., C. Delidakis, I. Livadaras, M. Tsagris, and M. Tabler. 2001. Short 5'-phosphorylated double-stranded RNAs induce RNA interference in *Drosophila*. Curr Biol 11:1776-1780.

- Bouwman, P., and S. Philipsen. 2002. Regulation of the activity of Sp1-related transcription factors. *Molecular and Cellular Endocrinology* 195:27-38.
- Boxshall, G. A. 2004. The evolution of arthropod limbs. *Biol. Rev. Cam. Phil. Soc.* 79:253-300.
- Brena, C., A. Chipman, A. Minelli, and M. Akam. 2006. Expression of trunk Hox genes in the centipede *Strigamia maritima*: sense and anti-sense transcripts. *Evo & Dev* 8:252-265.
- Bridge, D. M., N. A. Stover, and R. E. Steele. 2000. Expression of a novel receptor tyrosine kinase gene and a *paired*-like homeobox gene provides evidence of differences in patterning at the oral and aboral ends of hydra. *Dev Biol* 220:253-262.
- Briggs, D. E. G., and D. Collins. 1988. A Middle Cambrian chelicerate from Mount Stephen, British Columbia. *Palaeontology* 31:779-798.
- Brook, W. J., and S. M. Cohen. 1996. Antagonistic interactions between Wingless and Decapentaplegic responsible for Dorsal-Ventral pattern in the *Drosophila* leg. *Science* 273:1373-1377.
- Brox, A., L. Puellas, B. Ferreira, and L. Medina. 2003. Expression of genes GAD67 and *Distal-less-4* in the forebrain of *Xenopus laevis* confirms a common pattern in tetrapods. *J. Comp. Neurol.* 461:370-393.
- Brusca, G. J., and R. C. Brusca. 2003. Phylum Arthropoda: The Cheliceriformes. Pp. 653-698 in G. J. Brusca and R. C. Brusca, eds. *Invertebrates*. Sinauer Associates.
- Brusca, R. C., and G. J. Brusca. 1990. *Invertebrates*. Sinauer
- Brussels. 2006. *Solenidia*. Royal Belgian Institute of Natural Sciences, Brussels.
- Budd, G. 2001. Tardigrades as 'stem-group arthropods': the evidence from the Cambrian fauna. *Zool Anz* 240:265-279.
- Budd, G. 2002. A palaeontological solution to the arthropod head problem. *Nature* 417:271-275.
- Burke, A. C., C. E. Nelson, B. A. Morgan, and C. J. Tabin. 1995. *Hox* genes and the evolution of vertebrate axial morphology. *Development* 121:333-346.
- Bynum, E., and P. Porter. 2006. Spider mites on high plains corn. *Texas Corn IPM Manual*.
- Caldwell, M.W. and Lee, M.S.Y. 1997. A snake with legs from the marine Cretaceous of the Middle East. *Nature* 386: 705-709.
- Cameron, C. B., J. R. Garey, and B. J. Swalla. 2000. Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *PNAS* 97:4469-4474.
- Campbell, G. 2002. Distalization of the *Drosophila* leg by graded EGF-receptor activity. *Nature* 418:781-784.
- Campbell, G., and A. Tomlinson. 1998. The roles of homeobox genes *aristaleless* and *Distal-less* in patterning the legs and wings of *Drosophila*. *Dev* 125:4483-4493.
- Campbell, G., T. Weaver, and A. Tomlinson. 1993. Axis specification in the developing *Drosophila* appendage: the role of *wingless*, *decapentaplegic* and the homeobox gene *aristaleless*. *Cell* 74:1113-1123.
- Caracciolo, A., A. Di Gregorio, F. Aniello, R. Di Lauro, and M. Branno. 2000. Identification and developmental expression of three *Distal-less* homeobox containing genes in the ascidian *Ciona intestinalis*. *Mechanisms of Development* 99:173-176.
- Carroll, S. B. 2005. *Endless Forms Most Beautiful: The New Science of Evo Devo*. Norton
- Carroll, S. B., J. K. Grenier, and S. D. Weatherbee. 2001. *From DNA to Diversity: molecular genetics and the evolution of animal design*. Blackwell Science
- Cartwright, P., M. Dick, and L. W. Buss. 1993. HOM/Hox type homeoboxes in the chelicerate *Limulus polyphemus*. *Mol Phyl & Evol* 2:185-192.
- Casares, F., M. Calleja, and E. Sánchez-Herrero. 1996. Functional similarity in appendage specification by the *Ultrabithorax* and *abdominal-A* *Drosophila* HOX genes. *The EMBO Journal* 15:3934-3942.
- Castelli-Gair, J., S. Greig, G. Micklem, and M. Akam. 1994. Dissecting the temporal requirements for homeotic gene function. *Dev* 120:1983-1995.
- Chipman, A., W. Arthur, and A. Akam. 2004a. Early development and segment formation in the centipede, *Strigamia maritima* (Geophilomorpha). *Evo & Dev* 6:78-89.
- Chipman, A., W. Arthur, and M. Akam. 2004b. A double segment periodicity underlies segment generation in centipede development. *Curr. Biol.* 14:1250-1255.
- Clark, S. R. L. 2000. *Biology and Christian Ethics*. Cambridge University Press
- Cliffe, A., F. Hamada, and M. Bienz. 2003. A role of Dishevelled in relocating axin to the plasma membrane during wingless signaling. *Curr Biol* 13:960-966.
- Cloudsley-Thompson, J. L. 1958. Mites and Ticks. Pp. 195-203. *Spiders, Scorpions, Centipedes and Mites*. Pergamon Press.
- Coates, M. and Ruta, M. 2000. Nice snake, shame about the legs. *TREE* 15 (12): 503-507.

- Cobos, I., V. Broccoli, and J. L. Rubenstein. 2005. The vertebrate ortholog of *Aristaless* is regulated by *Dlx* genes in the developing forebrain. *J. Comp. Neurol.* 483:292-303.
- Coddington, J. A., G. Giribet, M. S. Harvey, L. Prendini, and D. E. Walter. 2004. Arachnida. Pp. 296-318 in J. Cracraft and M. J. Donoghue, eds. *Assembling the Tree of Life*. Oxford University Press.
- Cohen, B., A. A. Simcox, and S. Cohen. 1993. Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. *Dev* 117:597-608.
- Cohen, B., E. A. Wimmer, and S. M. Cohen. 1991. Early development of leg and wing primordia in the *Drosophila* embryo. *Mech. Dev.* 33:229-240.
- Cohen, S. M. 1990. Specification of limb development in the *Drosophila* embryo by positional cues from segmentation genes. *Nature* 343:173-177.
- Cohen, S. M., and G. Yurgens. 1989. Proximo-distal pattern formation in *Drosophila*: cell autonomous requirement for *Distal-less* gene activity. *EMBO J* 8:2045-2055.
- Cohn, M. J., and C. Tickle. 1999. Developmental basis of limblessness and axial patterning in snakes. *Nature* 399:474-479.
- Collyer, E. 1998. HortFACT: Two-spotted spider mite life cycle in T. H. a. F. R. I. o. N. Z. Ltd., ed. The Horticulture and Food Research Institute of New Zealand Ltd.
- Cook, C. E., E. Jiménez, M. Akam, and E. Saló. 2004. The Hox gene complement of acoel flatworms, a basal bilaterian clade. *Evo & Dev* 6:154-163.
- Cook, C. E., M. L. Smith, M. J. Telford, A. Bastianello, and M. Akam. 2001. *Hox* genes and the phylogeny of the arthropods. *Curr Biol* 11:759-763.
- Cooke, R. M., A. J. Wilkinson, M. Baron, A. Pastore, M. J. Tappin, I. D. Campbell, H. Gregory, and B. Sheard. 1987. The solution structure of human epidermal growth factor. *Nature* 327:339-341.
- Cornec, J.-P., and A. Gilles. 2006. Ubilateria, un être évolué? *Medecine/Sciences* 22:493-501.
- Couso, J. P., and A. Martinez Arias. 1994. *Notch* is required for *wingless* signaling in the epidermis of *Drosophila*. *Cell* 79:259-272.
- Couzin, J. 2002. Small RNAs make big splash. *Science* 298:2296-2297.
- Cracraft, J., and M. J. Donoghue. 2004. *Assembling the Tree of Life*. Oxford University Press
- Crooker, Allen, and Moffitt. 2003. Functional morphology of the spider mite gnathosoma.
- Dahn, R. D., and J. F. Fallon. 1999. Limbining outgrowth: BMPs as negative regulators in limb development. *BioEssays* 21:721-725.
- Damen, W. G. 2002. Parasegmental organisation of the spider embryo implies that the parasegment is an evolutionarily conserved entity in arthropod embryogenesis. *Dev* 129:1239-1250.
- Damen, W. G. M., M. Hausdorf, E.-A. Seyfarth, and D. Tautz. 1998. A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *PNAS* 95:10665-10670.
- Damen, W. G. M., R. Janssen, and N.-M. Prpic. 2005. Pair rule gene orthologs in spider segmentation. *Evo & Dev* 7:618-628.
- Damen, W. G. M., T. Saradaki, and M. Averof. 2002. Diverse adaptations of an ancestral gill: a common evolutionary origin for wings, breathing organs, and spinnarets. *Curr Biol* 12:1711-1716.
- Damen, W. G. M., M. Weller, and D. Tautz. 2000. Expression patterns of *hairy*, *even-skipped*, and *runt* in the spider *Cupiennius salei* imply that these genes were segmentation genes in a basal arthropod. *PNAS* 97:4515-4519.
- Darras, S., and H. Nishida. 2001. The BMP/Chordin antagonism controls sensory pigment cell specification and differentiation in the ascidian embryo. *Dev Biol* 236:271-288.
- Darwin, C. 1859. *The Origin of Species*. Oxford University Press
- Davidson, E. H. 2001. *Genomic Regulatory Systems: Development and Evolution*. Academic Press
- Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311:796-800.
- Davidson, E. H., D. McClay, and L. Hood. 2003. Regulatory gene networks and the properties of the developmental process. *PNAS* 100:1475-1480.
- Davis, G. K., and N. H. Patel. 2002. Short, long, and beyond: Molecular and embryological approaches to insect segmentation. *Ann Rev Entomol* 47:669-699.
- de Beer, G. R. 1940. *Embryos and Ancestors*. Oxford University Press
- de Celis, J. F., R. Barrio, A. d. Arco, and A. García-Bellido. 1993. Genetic and molecular characterisation of a Notch mutation in its Delta- and Serrate-binding domain in *Drosophila*. *PNAS* 90:4037-4041.
- De Robertis, E. M., and Y. Sasai. 1996. A common plan for dorsoventral patterning in Bilateria. *Nature*

- 380:37-40.
- de Rosa, R., J. K. Grenier, T. Andreeva, C. E. Cook, A. Adoutte, M. Akam, S. Carroll, and G. Balavoine. 1999. Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399:772-776.
- Dealy, C. N., V. Scranton, and H.-C. Cheng. 1998. Roles of transforming growth factor- α and epidermal growth factor in chick limb development. *Dev Biol* 202:43-55.
- Dearden, P., M. Grbic, and C. Donly. 2003. Vasa expression and germ cell specification in the spider mite *Tetranychus urticae*. *Dev Genes Evol* 212:599-603.
- Dearden, P. K., and M. Akam. 2001. Early embryo patterning in the grasshopper, *Schistocerca gregaria*: *wingless*, *decapentaplegic* and *caudal* expression. *Dev* 128:3435-3444.
- Dearden, P. K., C. Donly, and M. Grbic. 2002. Expression of pair-rule gene homologues in a chelicerate: early patterning of the two-spotted spider mite *Tetranychus urticae*. *Development* 129:5461-5472.
- Decotto, E., and E. L. Ferguson. 2001. A positive role for Short gastrulation in modulating BMP signaling during dorsoventral patterning in the *Drosophila* embryo. *Development* 128:3831-3841.
- Delfini, M.-C., J. Dubrulle, P. Malapert, J. Chal, and O. Pourquié. 2005. Control of the segmentation process by graded MAPK/ERK activation in the chick embryo. *PNAS* 102:11343-11348.
- Delsuc, F., H. Brinkmann, D. Chourrout, and H. Philippe. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965-968.
- Deutsch, J. S. 2004. Segments and parasegments in Arthropods: a functional perspective. *BioEssays* 26:1117-1125.
- Deutsch, J. S., and E. Mouchel-Vielh. 2003. Hox genes and the crustacean body plan. *BioEssays* 25.9:878-887.
- Dhawan, S., and K. P. Gopinathan. 2003. Spatio-temporal expression of *wnt-1* during embryonic-, wing- and silk gland development in *Bombyx mori*. *Gene Expression Patterns* 3:559-570.
- Diaz-Benjumea, F. J., B. Cohen, and S. B. Cohen. 1994. Cell interaction between compartments establishes the proximo-distal axis of *Drosophila* legs. *Nature* 372:175-178.
- DiNardo, S., J. M. Kuner, J. Theis, and P. H. O'Farrell. 1985. Development of embryonic pattern in *D. melanogaster* as revealed by accumulation of the nuclear *engrailed* pattern. *Cell* 43:59-69.
- Drossopoulou, G., K. E. Lewis, J. J. Sanz-Esquerro, N. Nikbakht, A. P. McMahon, and C. Hofmann. 2000. A model for anteroposterior patterning of the vertebrate limb based on sequential long- and short-range Shh signalling and Bmp signalling. *Development* 127:1337-1348.
- Du, H., R. A. Fuh, J. Li, A. Corkan, and J. S. Lindsey. 1998. PhotochemCAD: A computer-aided design and research tool in photochemistry. *Photochemistry and Photobiology* 68:141-142.
- Dubrulle, J., M. J. McGrew, and O. Pourquié. 2001. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. *Cell* 106:219-232.
- Dudley, A. T., M. A. Ros, and C. J. Tabin. 2002. A re-examination of proximodistal patterning during vertebrate limb development. *Nature* 418:539-544.
- Duman-Scheel, M., and N. H. Patel. 1999. Analysis of molecular marker expression reveals neuronal homology in distantly related arthropods. *Dev* 126:2327-2334.
- Duman-Scheel, M., N. Pirkel, and N. H. Patel. 2002. Analysis of the expression pattern of *Mysidium columbiae* *wingless* provides evidence for conserved mesodermal and retinal patterning processes among insects and crustaceans. *Dev Genes Evol* 212:114-123.
- Dumstrei, K., C. Nassif, G. Abboud, A. Aryai, A. Aryai, and V. Hartenstein. 1998. EGFR signaling is required for the differentiation and maintenance of neural progenitors along the dorsal midline of the *Drosophila* embryonic head. *Development* 125:3417-3426.
- Dumstrei, K., F. Wang, and e. al. 2002. Interaction between EGFR signaling and DE-cadherin during nervous system morphogenesis. *Development* 129:3983-3994.
- Dunlop, J. A., and C. P. Arango. 2005. Pycnogonid affinities: a review. *JZS* 43:8-21.
- Eisenberg, L. M., P. W. Ingham, and A. M. Brown. 1992. Cloning and characterization of a novel *Drosophila* *Wnt* gene, *Dwnt-5*, a putative downstream target of the homeobox gene *Distal-less*. *Dev Biol* 154:73-83.
- Elbashir, S. M., J. Martinez, A. Patkaniowska, W. Lendeckel, and T. Tuschl. 2001. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *The EMBO Journal* 20:6877-6888.
- Entchev, E. V., A. Schwabedissen, and M. González-Gaitán. 2000. Gradient formation of the TGF- β homolog Dpp. *Cell* 103:981-991.
- Erlich, H. A. 1985. PCR Technology: Principles and Applications for DNA Amplification. Stockton Press

- Estella, C., G. Rieckhof, M. Calleja, and G. Morata. 2003. The role of buttonhead and Sp1 in the development of the ventral imaginal discs of *Drosophila*. *Dev* 130:5929-5941.
- Ewing, H. E. 1928. The legs and leg-bearing segments of some primitive arthropod groups, with notes on leg-segmentation in the Arachnida. *Smithsonian Miscellaneous Collections* 80:1-41.
- Extavour, C. 2003. Phosphorylated Histone-3 antibody stability *in C. Cyrus-Kent*, ed.
- Farley, R. D. 1998. Matrotrophic adaptations and early stages of embryogenesis in the desert scorpion *Paruroctonus mesaensis* (Vaejovidae). *J. Morphol.* 237:187-211.
- Farley, R. D. 2001. Development of segments and appendages in embryos of the desert scorpion *Paruroctonus mesaensis* (Scorpiones: Vaejovidae). *J. Morphol.* 250:70-88.
- Farley, R. D. 2005. Developmental changes in the embryo, pronymph, and first moult of the scorpion *Centruroides vittatus* (Scorpiones: Buthidae). *J. Morphol.* 265:1-27.
- Feinberg, E. H., and C. P. Hunter. 2003. Transport of dsRNA into cells by the transmembrane protein SID-1. *Science* 301:1545-1547.
- Felsenstein, J. 2001. PHYLIP: Phylogeny Inference Package. University of Washington, Seattle, Washington.
- Finnerty, J. R., and M. Q. Martindale. 1997. Homeoboxes in sea anemones (Cnidaria; Anthozoa): a PCR-based survey of *Nematostella vectensis* and *Metridium senile*. *Biol. Bull.* 193:62-76.
- Finnerty, J. R., K. Pang, P. Burton, D. Paulson, and M. Q. Martindale. 2004. Origins of bilateral symmetry: *Hox* and *Dpp* expression in a sea anemone. *Science* 304:1335-1337.
- Fiol, C. J., A. M. Mahrenholz, Y. Wang, R. W. Roeske, and P. J. Roach. 1987. Formation of protein kinase recognition sites by covalent modification of the substrate. *J Biol Chem* 262:14042-14048.
- Fire, A., S. Q. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.
- Flechtmann, C. H. W., and D. K. Knihinicki. 2002. New species and new record of Tetranychus Dufour from Australia, with a key to the major groups in this genus based on females (Acari: Prostigmata: Tetranychidae). *Australian J Entomol* 41:118-127.
- Fleig, R. 1990. *Engrailed* expression and body segmentation in the honeybee *Apis mellifera*. *Roux's Arch Dev Biol* 198: 467-473.
- Fleming, R. J., T. N. Scottgale, R. J. Dieterich, and S. Artavanis-Tsakonas. 1990. The gene *Serrate* encodes a putative EGF-like transmembrane protein essential for proper ectodermal development in *Drosophila melanogaster*. *Genes & Dev* 4:2188-2201.
- Florida. 2005. Two spotted spider mite: *Tetranychus urticae* Koch. University of Florida.
- Foelix, R. F. 1996. *Biology of Spiders*. Oxford University Press/Georg Thieme Verlag
- Fondon, J. W., and H. R. Garner. 2004. Molecular origins of rapid and continuous morphological evolution. *PNAS* 101:18058-18063.
- Fujise, M., S. Takeo, and e. al. 2003. Dally regulates Dpp morphogen gradient formation in the *Drosophila* wing. *Dev* 130:1515-1522.
- Gabay, L., R. Seger, and B.-Z. Shilo. 1997. In situ activation pattern of *Drosophila* EGF receptor pathway during development. *Science* 277:1103-1106.
- Galant, R., and S. B. Carroll. 2002. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415:910-913.
- Galant, R., C. M. Walsh, and S. B. Carroll. 2002. Hox repression of a target gene: extra-denticle independent, additive action through multiple monomer binding sites. *Dev* 129:3115-3126.
- Galindo, M. I., S. A. Bishop, S. Greig, and J. P. Couso. 2002. Leg patterning driven by proximal-distal interactions and EGFR signaling. *Science* 297:256-259.
- Gallitano-Mendel, A., and R. Finkelstein. 1998. Ectopic *orthodenticle* expression alters segment polarity gene expression but not head segment identity in the *Drosophila* embryo. *Dev Biol* 199:125-137.
- Garey, J. R., M. Krotec, D. R. Nelson, and J. Brooks. 1996. Molecular analysis supports a tardigrade-arthropod association. *Invertebrate Biology* 115:79-88.
- Gebelein, B., J. Culi, H. D. Ryoo, W. Zhang, and R. S. Mann. 2002. Specificity of *Distalless* repression and limb primordia development by abdominal Hox proteins. *Dev Cell* 3:487-498.
- Gebelein, B., D. J. McKay, and R. S. Mann. 2004. Direct integration of Hox and segmentation gene inputs during *Drosophila* development. *Nature* 431:653-659.
- Gellner, K., and S. Brenner. 1999. Analysis of 148 kb of genomic DNA around the *wnt1* locus of *Fugu rubripes*. *Genome Research* 9:251-258.
- Gerlitz, O., and K. Basler. 2002. Wingful, an extracellular feedback inhibitor of Wingless. *Genes & Dev*

- 16:1055-1059.
- Gibert, J.-M., E. Mouchel-Vielh, E. Quéinnec, and J. S. Deutsch. 2000. Barnacle duplicate *engrailed* genes: divergent expression patterns and evidence for a vestigial abdomen. *Evo & Dev* 2:194-202.
- Gilbert, S. F. 2000. *Developmental Biology*. Sinauer Associates Inc.
- Giorgianni, M. W., and N. H. Patel. 2004. Patterning of the branched head appendages in *Schistocerca americana* and *Tribolium castaneum*. *Evo & Dev* 6:402-410.
- Giribet, G. 2002. Current advances in the phylogenetic reconstruction of metazoan evolution: A new paradigm for the Cambrian explosion? *Mol. Phyl. & Evol.* 24:345-357.
- Giribet, G., S. Carranza, J. Baguña, M. Ruitort, and C. Ribera. 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Mol Biol Evol* 13:76-84.
- Giribet, G., G. D. Edgecombe, and W. C. Wheeler. 1999. Sistemática y filiogenia de artrópodos: estado de la cuestión con énfasis en análisis de datos moleculares. *Bol. S.E.A.* 26:197-212.
- Gordon, M. D., and R. Nusse. 2006. Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors. *J Biol Chem* 281:22429-22433.
- Goto, S., and S. Hayashi. 1997. Specification of the embryonic limb primordium by graded activity of Decapentaplegic. *Dev* 124:125-132.
- Grandjean, F. 1935. Les poils et les organes sensitifs portés par les pattes et le palpe chez les Oribates (1e partie). *Bull. Soc. Zool. France* 60:6-39.
- Grandjean, F. 1936. Observations sur les Acariens (3e série). *Bull. Mus. Natl. Hist. Nat. Paris Ser. 2* 8:84-91.
- Grandjean, F. 1939. Observations sur les Acariens (5e série). *Bull. Mus. Natl. Hist. Nat. Paris Ser. 2* 11:393-401.
- Grandjean, F. 1948. Quelques caractères des Tétranyques. *Bull. Mus. Natl. Hist. Nat. Paris Ser. 2* 20:517-524.
- Grbic, M. 2000. "Alien" wasps and evolution of development. *BioEssays* 22:920-932.
- Grbic, M., Nagy, L.M., Carroll, S.B., Strand, M. 1996. Polyembryonic development: insect pattern formation in a cellularised environment. *Dev* 122: 795-804.
- Green, F. J. 1990. 4',6-Diamidino-2-phenylindole (DAPI). Pp. 244. Sigma-Aldrich Handbook of Stains, Dyes and Indicators.
- Greenwald, I. 1985. *lin-12*, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. *Cell* 43:583-590.
- Grenier, J. K., and S. B. Carroll. 2000. Functional evolution of the Ultrabithorax protein. *PNAS* 97:704-709.
- Grenier, J. K., T. L. Garber, R. Warren, P. M. Whittington, and S. Carroll. 1997. Evolution of the entire arthropod *Hox* gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr Biol* 7:547-553.
- Grienerberger, A., S. Merabet, J. Manak, and e. al. 2003. Tgf-beta signaling acts on a Hox response element to confer specificity and diversity to Hox protein function. *Dev* 130:5445-5455.
- Grishok, A., A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, and C. C. Mello. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106:23-34.
- Greene, H.W. and Cundall, D. 2000. Limbless tetrapods and snakes with legs. *Science* 287: 1937-1941.
- Gullan, P. J., and P. S. Cranston. 2000. *The Insects: an outline of entomology*. Blackwell Science
- Haas, M. S., S. J. Brown, and R. W. Beeman. 2001a. Homeotic evidence for the appendicular origin of the labrum in *Tribolium castaneum*. *Dev Genes Evol* 211:96-102.
- Haas, M. S., S. J. Brown, and R. W. Beeman. 2001b. Pondering the procephalon: the segmental origin of the labrum. *Dev Genes Evol* 211:89-95.
- Hammond, S. M., A. A. Caudy, and G. J. Hannon. 2001. Post-transcriptional gene silencing by double-stranded RNA. *Nature Revs Genet* 2:110-119.
- Harzsch, S. 2003. Development of the arthropod nervous system: variations on a common theme? *Arth. Str. & Dev.* 32:3-4.
- Hayward, D. C., G. Samuel, P. C. Pontynen, J. Catmull, R. Saint, D. J. Miller, and E. E. Ball. 2002. Localised expression of a *dp/BMP2/4* ortholog in a coral embryo. *PNAS* 99:8106-8111.
- Hayward, P., K. Brennan, P. Sanders, T. Balayo, R. DasGupta, N. Perrimon, and A. Martinez Arias. 2005. Notch modulates Wnt signalling by associating with Armadillo/beta-catenin and regulating its transcriptional activity. *Development* 132:1819-1830.
- Hejnal, A., and G. Scholtz. 2004. Clonal analysis of *Distal-less* and *engrailed* expression patterns during early morphogenesis of uniramous and biramous crustacean limbs. *Dev Genes Evol* 214:473-485.

- Hemmati-Brivanlou, A., J. R. de la Torre, C. Holt, and R. M. Harland. 1991. Cephalic expression and molecular characterisation of *Xenopus En-2*. *Development* 111:715-724.
- Henry, C. M. 2003. High hopes for RNA interference. *C & EN, Science & Technology*:32-36.
- Henry, C. M. 2004. Another piece of the puzzle: messenger RNA-cleaving protein, active in RNA interference, is identified as Argonaute2. *Chem. Eng. News* Aug 9:30.
- Hirth, F., L. Kammermeier, E. Frei, U. Walldorf, M. Noll, and H. Reichert. 2003. An urbilaterian origin for the tripartite brain: developmental insights from *Drosophila*. *Dev* 130:2365-2373.
- Hobmayer, B., F. Rentzsch, K. Kuhn, C. M. Happel, C. C. von Laue, P. Snyder, U. Rothbacher, and T. W. Holstein. 2000. WNT signalling molecules act in axis formation in the diploblastic metazoan *Hydra*. *Nature* 407:186-189.
- Hoffmann, F. M., L. D. Fresco, H. Hoffmann-Falk, and B.-Z. Shilo. 1983. Nucleotide sequences of the *Drosophila src* and *abl* homologs: conservation and variability in the *src* family oncogenes. *Cell* 35:393-401.
- Holland, P. 2004. The ups and downs of a sea anemone. *Science* 304:1255-1256.
- Holland, P. W. H., and N. A. Williams. 1990. Conservation of *engrailed*-like homeobox sequences during vertebrate evolution. *FEBS Letts.* 277:250-252.
- Holm, A. 1952. Experimentelle untersuchungen über die entwicklung und entwicklungphysiologie des spinnenembryos. *Zool. Bidr. Uppsala* 29:293-424.
- Hsia, C. C., and W. McGinnis. 2003. Evolution of transcription factor function. *Curr Opin* 13:199-206.
- Hsuan, J. J., G. Panayotou, and M. D. Waterfield. 1989. Structural basis for epidermal growth factor receptor function. *Progress in Growth Factor Research* 1:23-32.
- Hueber, S. D., D. Bezdán, S. R. Henz, M. Blank, H. Wu, and I. Lohmann. 2007. Comparative analysis of Hox downstream genes in *Drosophila*. *Development* 134:381-392.
- Hughes, C. L., and T. C. Kaufman. 2001. Exploring the myriapod body plan: expression patterns of ten Hox genes in a centipede. *Dev* 129:1225-1238.
- Hughes, C. L., and T. C. Kaufman. 2002a. Exploring myriapod segmentation: the expression patterns of even-skipped, engrailed, and wingless in a centipede. *Dev Biol* 247:47-61.
- Hughes, C. L., and T. C. Kaufman. 2002b. Hox genes and the evolution of the arthropod body plan. *Evo & Dev* 4:459-499.
- Hwang, U. W., M. Friedrich, D. Tautz, C. J. Park, and K. Woo. 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* 413:154-157.
- Imai, K. S. 2003. Isolation and characterisation of beta-catenin downstream genes in early embryos of the ascidian *Ciona savignyi*. *Differentiation* 71:346-360.
- Ingham, P. W., and A. Martinez Arias. 1992. Boundaries and fields in early embryos. *Cell* 68:221-235.
- Ip, Y. T., R. E. Park, D. Kosman, E. Bier, and M. Levine. 1992. The dorsal gradient morphogen regulates stripes of *rhomboid* expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes & Dev* 6:1728-1739.
- Irish, V. F., and W. M. Gelbart. 1987. The *decapentaplegic* gene is required for dorso-ventral patterning of the *Drosophila* embryo. *Genes & Dev* 1:868-879.
- Itow, T., S. Kenmochi, and T. Mochizuki. 1991. Induction of secondary embryos by intra- and interspecific grafts of center cells under the blastopore in horseshoe crabs. *Dev, Growth & Differ.* 33:251-258.
- Jaffe, L., H. D. Ryoo, and R. S. Mann. 1997. A role for phosphorylation by casein kinase II in modulating Antennapedia activity in *Drosophila*. *Genes & Dev* 11:1327-1340.
- Janssen, R. 2005. Myriapod Sp genes, pers. com.
- Janssen, R., N.-M. Prpic, and W. G. M. Damen. 2004. Gene expression suggests decoupled dorsal and ventral segmentation in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). *Dev Biol* 268:89-104.
- Jaynes, J. B., and M. Fujioka. 2004. Drawing lines in the sand: *even skipped* et al. and parasegment boundaries. *Dev Biol* 269:609-622.
- Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23:403-405.
- Jenner, R. A. 2004. The scientific status of metazoan cladistics: why current research practice must change. *Zoologica Scripta* 33:293-310.
- Jeppson, L. R., H. H. Keifer, and E. W. Baker. 1975. Mites Injurious to Economic Plants. University of California Press, Berkley, USA.

- Jia, J., K. Amanai, G. Wang, J. Tang, B. Wang, and J. Jiang. 2002. Shaggy/GSK3 antagonises Hedgehog signalling by regulating Cubitus interruptus. Nature advance online publication DOI 10.1031/nature733
- Jiang, D., and W. C. Smith. 2002. An ascidian *engrailed* gene. Dev Genes Evol 212:399-402.
- Jin, M., K. Sawamoto, M. Ito, and H. Okano. 2000. The interaction between the *Drosophila* secreted protein Argos and the Epidermal Growth Factor Receptor inhibits dimerization of the receptor and binding of secreted Spitz to the receptor. Mol Cell Biol 20:2098-2107.
- Jockusch, E. L., C. Nulsen, S. J. Newfield, and L. M. Nagy. 2000. Leg development in flies versus grasshoppers: differences in dpp expression do not lead to differences in the expression of downstream components of the leg patterning pathway. Dev 127:1617-1626.
- Jockusch, E. L., and K. A. Ober. 2004. Hypothesis testing in evolutionary developmental biology: a case study from insect wings. J. Heredity 95:382-396.
- Jockusch, E. L., T. A. Williams, and L. M. Nagy. 2004. The evolution of patterning of serially homologous appendages in insects. Dev Genes Evol 214:324-338.
- Johnson, K. R., J. E. Lewis, D. Li, J. Wahl, A. Peralta Soler, K. A. Knudsen, and M. J. Wheelock. 1993. P- and E-cadherin are in separate complexes in cells expressing both cadherins. Exp. Cell Res. 207:252-260.
- Jones, S. E., and C. Jomary. 2002. Secreted Frizzled-related proteins: searching for relationships and patterns. BioEssays 24:811-820.
- Jorissen, R. N., F. Walker, N. Pouliot, T. P. J. Garrett, C. W. Ward, and A. W. Burgess. 2003. Epidermal growth factor receptor: mechanisms of activation and signalling. Exp. Cell Res. 284:31-53.
- Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51:1079-1090.
- Kamm, K., B. Schierwater, W. Jakob, S. L. Dellaporta, and D. J. Miller. 2006. Axial patterning and diversification in the Cnidaria predate the Hox system. Curr. Biol. 16:920-926.
- Karr, T. L., M. P. Weir, Z. Ali, and T. B. Kornberg. 1989. Patterns of *engrailed* protein in early *Drosophila* embryos. Development 105:605-612.
- Kawakami, Y., C. Rodríguez Esteban, and e. al. 2004. Sp8 and Sp9, two closely related buttonhead-like transcription factors, regulate Fgf8 expression and limb outgrowth in vertebrate embryos. Development 131:4763-4774.
- Kearney, M. and Stuart, B.L. 2004. Repeated evolution of limblessness and digging heads in worm lizards revealed by DNA from old bones. Proc. R. Soc. Lon. B 271: 1677-1683.
- Kelsh, R., R. O. J. Weinzierl, R. A. H. White, and M. Akam. 1994. Homeotic gene expression in the locust *Schistocerca*: an antibody that detects conserved epitopes in *Ultrabithorax* and *Abdominal-A* proteins. Dev. Genet. 15:19-31.
- Kennedy, S., D. Wang, and G. Ruvkun. 2004. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. Nature 427:645-649.
- Kent, O. A., and A. M. MacMillan. 2004. RNAi: P running interference for the cell. Org. Biomol. Chem. 2:1957-1961.
- Kettle, C., J. Johnstone, T. Jowett, H. Arthur, and W. Arthur. 2003. The pattern of segment formation, as revealed by *engrailed* expression, in a centipede with a variable number of segments. Evo & Dev 5:198-207.
- Khila, A. 2006a. Antibodies cross-reactive in *Tetranychus urticae*, pers. com.
- Khila, A. 2006b. Feasibility of *Dll* RNAi in *Tetranychus urticae*, pers. com.
- Kim, S. H., and S. T. Crews. 1993. Influence of *Drosophila* ventral epidermal development by the CNS midline cells and *spitz* class genes. Development 118:893-901.
- King, M. C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107-116.
- King, N., and S. Carroll. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. PNAS 98:15032-15037.
- Kingsley, D. M. 1994. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes & Dev 8:133-146.
- Klämbt, C. 2001. EGF receptor signalling: roles of Star and Rhomboid revealed. Curr Biol 12: r21-r23.
- Klein, D. E., V. M. Nappi, and e. al. 2004. Argos inhibits epidermal growth factor receptor signalling by ligand sequestration. Nature 430:1040-1044.

- Kobayashi, M., M. Fujioka, E. N. Tolkunova, D. Deka, M. Abu-Shaar, R. S. Mann, and J. B. Jaynes. 2003. Engrailed cooperates with extradentical and homothorax to repress target genes in *Drosophila*. *Dev* 130:741-751.
- Koch, C. L. 1836. *Tetranychus urticae* in F. Pustet, ed. Deutschlands Crustaceen, Myriapoden und Arachniden: Ein Beitrag zur Deutschen Fauna. Regensburg.
- Kraft, R., and H. Jäckle. 1994. *Drosophila* mode of metamerisation in the embryogenesis of the lepidopteran insect *Manduca sexta*. *PNAS* 91:6634-6638.
- Krantz, G. W., and E. E. Lindquist. 1979. Evolution of phytophagous mites (Acari). *Ann Rev Entomol* 24:121-158.
- Krebs, E. G., and J. A. Beavo. 1979. Phosphorylation-dephosphorylation of enzymes. *Ann Rev Biochem* 48:923-959.
- Kretzschmar, M., J. Doody, I. Timokhina, and J. Massagué. 1999. A mechanism of repression of TGF-beta/Smad signaling by oncogenic Ras. *Genes & Dev* 13:804-816.
- Kubota, K., S. Goto, K. Eto, and S. Hayashi. 2000. EGF receptor attenuates Dpp signaling and helps to distinguish the wing and leg cell fates in *Drosophila*. *Dev* 127:3769-3776.
- Kubota, K., S. Goto, and S. Hayashi. 2003. The role of Wg signaling in the patterning of embryonic leg primordium in *Drosophila*. *Dev Biol* 257:117-126.
- Kumar, J. P., and K. Moses. 2001. EGF receptor and Notch signaling act upstream of *eyeless/Pax6* to control eye specification. *Cell* 104:687-697.
- Kumar, J. P., M. Tio, F. Hsiung, S. Akopyan, L. Gabay, R. Seger, and B.-Z. Shilo. 1998. Dissecting the roles of the *Drosophila* EGF receptor in eye development and MAP kinase activation. *Development* 125:3875-3885.
- Kusche, K., and T. Burmester. 2001. Diplopod hemocyanin sequence and the phylogenetic position of the myriapoda. *Mol Biol Evol* 18:1566-1573.
- Kusserov, A., K. Pang, C. Sturm, M. Hrouda, J. Lentfer, H. A. Schmidt, U. Technau, A. von Haeseler, B. Hobmayer, M. Q. Martindale, and T. W. Holstein. 2005. Unexpected complexity of the *Wnt* gene family in a sea anemone. *Nature* 433:156-160.
- Lai, E. C. 2004. Notch signaling: control of cell communication and cell fate. *Development* 131:965-973.
- Lallemand, Y., M.-A. Nicola, C. Ramos, A. Bach, C. Saint Clément, and B. Robert. 2005. Analysis of *Msx1;Msx2* double mutants reveals multiple roles for *Msx* genes in limb development. *Development* 132:3003-3014.
- Lam, G., and C. S. Thummel. 2000. Inducible expression of double-stranded RNA directs specific genetic interference in *Drosophila*. *Curr Biol* 10:957-963.
- Laney, J. D., and M. D. Biggin. 1996. Redundant control of *Ultrabithorax* by *zeste* involves functional levels of *zeste* protein binding at the *Ultrabithorax* promoter. *Development* 122:2303-2311.
- Larsen, C. W., E. Hirst, C. Alexandre, and J.-P. Vincent. 2003. Segment boundary formation in *Drosophila* embryos. *Dev* 130:5625-5635.
- Laurie, M. 1890. The embryology of the scorpion *Euscorpium italicus*. *Quart J Micr Sci* 31:105-141.
- Lax, I., A. Johnson, R. Howk, J. Sap, F. Bellot, M. Winkler, A. Ullrich, B. Vennstrom, J. Schlessinger, and D. Givol. 1988. Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mouse cells, and differential binding of EGF and transforming growth factor alpha. *Molecular and Cellular Biology* 8:1970-1978.
- Lecuit, T., and S. B. Cohen. 1997. Proximo-distal axis formation in the *Drosophila* leg. *Nature* 388:139-145.
- Lee, M.S.Y., Scanlon, J.D. et al. 2000. Snake origins. *Science* 288: 1343-1345.
- Lee, S. E., R. D. Gates, and D. K. Jacobs. 2001. The isolation of a *Distal-less* gene fragment from two molluscs. *Dev Genes Evol* 211:506-508.
- Lee, S. E., and D. K. Jacobs. 1999. Expression of *Distal-less* in molluscan eggs, embryos, and larvae. *Evo & Dev* 1:172-179.
- Lemons, D., and W. McGinnis. 2006. Genomic evolution of *Hox* gene clusters. *Science* 313:1918-1922.
- Lepage, T., C. Ghiglione, and C. Gache. 1992. Spatial and temporal expression pattern during sea urchin embryogenesis of a gene coding for a protease homologous to the human protein BMP1 and to the product of the *Drosophila* dorsal-ventral patterning gene *tolloid*. *Development* 114:147-163.
- Levine, M. 2002. How insects lose their limbs. *Nature* 415:848-849.
- Lewin, B. 2002. *Genes VII*. OUP
- Lewis, D. L., M. DeCamillis, and R. L. Bennett. 2000. Distinct roles of the homeotic genes *Ubx* and *abd-A* in

- beetle embryonic abdominal appendage development. PNAS USA 97:4504-4509.
- Lewis, E. B. 1963. Genes and developmental pathways. Amer. Zool. 3:33-56.
- Lipardi, C., Q. Wei, and B. M. Paterson. 2001. RNAi and random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. Cell 107:297-307.
- Livneh, E., L. Glazer, D. Segal, J. Schlessinger, and B.-Z. Shilo. 1985. The *Drosophila* EGF receptor gene homolog: conservation of both hormone binding and kinase domains. Cell 40:599-607.
- Llimargas, M., and P. A. Lawrence. 2001. Seven Wnt homologues in *Drosophila*: a case study of the developing tracheae. PNAS 98:14487-14492.
- Logan, M. 2003. Finger or toe: the molecular basis of limb identity. Development 130:6401-6410.
- Lohmann, I. 2006. *Hox* genes: realising the importance of realisers. Curr. Biol. 16:R988-R989.
- Lowe, C. J., and G. A. Wray. 1997. Radical alterations in the roles of homeobox genes during echinoderm evolution. Nature 389:718-721.
- Lowe, C. J., M. Wu, A. Salic, L. Evans, E. Lander, N. Stange-Thomann, C. E. Gruber, J. Gerhart, and M. Kirschner. 2003. Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. Cell 113:853-865.
- Lycett, G., C. Blass, and C. Louis. 2001. Developmental variation in epidermal growth factor receptor size and localisation in the malaria mosquito, *Anopheles gambiae*. Insect Molecular Biology 10:619-628.
- Maddison, D. R., and W. P. Maddison. 2000. MacClade 4. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Mann, R. S., and S. B. Carroll. 2002. Molecular mechanisms of selector gene function and evolution. Curr. Opin. Genes & Dev. 12:592-600.
- Martinez Arias, A., and A. Stewart. 2002. Molecular Principles of Animal Development. Oxford University Press
- Martinez, P., J. C. Lee, and E. H. Davidson. 1997. Complete sequence of *SpHox8* and its linkage in the single *Hox* gene cluster of *Strongylocentrotus purpuratus*. J Mol Evol 44:371-377.
- Matilla, E., T. Pellinen, J. Nevo, K. Vuoriluoto, A. Arjonen, and J. Ivaska. 2005. Negative regulation of EGFR signaling through integrin- $\alpha_1\beta_1$ -mediated activation of protein tyrosine phosphatase TCPTP. Nature Cell Biology 7:78-85.
- Matzke, M. A., and J. A. Birchler. 2005. RNAi-mediated pathways in the nucleus. Nature Reviews Genetics 6:24-35.
- Mayer, U., and C. Nusslein-Volhard. 1988. A group of genes required for pattern formation in the ventral ectoderm of the *Drosophila* embryo. Genes & Dev 2:1496-1511.
- McMahon, J. A., and A. P. McMahon. 1989. Nucleotide sequence, chromosomal localization and developmental expression of the mouse *int-1*-related gene. Development 107:643-650.
- Merabet, S., Z. Kambris, and e. al. 2003. The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. Developmental Cell 4:761-768.
- Merlo, G. R., B. Zerega, L. Paleari, S. Trombino, S. Mantero, and G. Levi. 2000. Multiple functions of *Dlx* genes. Int J Dev Biol 44:619-626.
- Minelli, A. 2002. Homology, limbs and genitalia. Evo & Dev 4:127-132.
- Minnesota. 2006. Acari: Psoroptidae. University of Minnesota.
- Mittmann, B. 2002. Early neurogenesis in the horseshoe crab *Limulus polyphemus* and its implication for arthropod relationships. Biol Bull 203:221-222.
- Mittmann, B., and G. Scholtz. 2001. Distal-less expression in embryos of *Limulus polyphemus* (Chelicerata, Xiphosura) and *Lepisma saccharina* (Insecta, Zygentoma) suggests a role in the development of mechanoreceptors, chemoreceptors and the CNS. Dev Genes Evol 211:232-243.
- Mittmann, B., and G. Scholtz. 2003. Development of the nervous system in the "head" of *Limulus polyphemus* (Chelicerata: Xiphosura): morphological evidence for a correspondence between the segments of the chelicerae and of the (first) antennae of Mandibulata. Dev Genes Evol 213:9-17.
- Miyawaki, K., T. Mito, I. Sarashina, H. Zhang, Y. Shinmyo, H. Ohuchi, and S. Noji. 2004. Involvement of Wingless/Armadillo signaling in the posterior sequential segmentation in the cricket, *Gryllus bimaculatus* (Orthoptera), as revealed by RNAi analysis. Mechanisms of Development 121:119-130.
- Momose, T., and E. Houliston. 2007. Two oppositely localised *frizzled* RNAs as axis determinants in a cnidarian embryo. PLoS Biology 5:1-11.
- Monnier, V., K. S. Ho, M. Sanial, M. P. Scott, and A. Plessis. 2002. Hedgehog signal transduction proteins: contacts of the Fused kinase and Ci transcription factor with the kinesin-related protein Costal2. Dev

- Moon, R. T., J. D. Brown, J. A. Yang-Snyder, and J. R. Miller. 1997. Structurally related receptors and antagonists compete for secreted Wnt ligands. *Cell* 88:725-728.
- Moore, J. 2001. An Introduction to the Invertebrates. CUP
- Morata, G., and E. Sánchez-Herrero. 1999. Patterning mechanisms in the body trunk and the appendages of *Drosophila*. *Dev* 126:2823-2828.
- Mouchel-Vielh, E., M. Blin, C. Rigolot, and J. S. Deutsch. 2002. Expression of a homologue of the *fushi-tarazu* (*ftz*) gene in a cirripede crustacean. *Evo & Dev* 4:76-85.
- Mouchel-Vielh, E., C. Rigolot, J.-M. Gibert, and J. S. Deutsch. 1998. Molecules and the body plan: the Hox genes of cirripedes (Crustacea). *Mol Phyl Evol* 9:382-389.
- Nambu, J. R., J. O. Lewis, K. A. Wharton, and S. T. Crews. 1991. The *Drosophila single-minded* gene encodes a Helix-Loop-Helix protein that acts as a master regulator of CNS midline development. *Cell* 67:1157-1167.
- Nardi, F., G. Spinsanti, and J. L. Boore. 2003a. Hexapod origins: monophyletic or paraphyletic? *Science* 299:1887-1889.
- Nardi, F., G. Spinsanti, J. L. Boore, A. Carapelli, R. Dallai, and F. Frati. 2003b. Response to comment on "Hexapod origins: monophyletic or paraphyletic?" *Science* 301:1482e.
- Návia, D. and Flechtmann, C.H.W. 2004. Rediscovery and redescription of *Tetranychus gigas* (Acari, Prostigmata, Tetranychidae). *Zootaxa* 547: 1-8.
- Nederbragt, A. J., A. E. van Loon, and W. J. A. G. Dictus. 2002. Expression of *Patella vulgata* orthologs of *engrailed* and *ppp-BMP2/4* in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. *Dev Biol* 246:341-355.
- Negrisol, E., A. Minelli, and G. Valle. 2004. The mitochondrial genome of the house centipede *Scutigera* and the monophyly versus paraphyly of myriapods. *Mol. Biol. Evol.* 21:770-780.
- Niedert, A. H., V. Virupannavar, G. W. Hooker, and J. A. Langeland. 2001. Lamprey *Dlx* genes and early vertebrate evolution. *PNAS* 96:1665-1670.
- Nishikura, K. 2001. A short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. *Cell* 107:415-418.
- Niwa, N., Y. Inoue, A. Nozawa, M. Saito, Y. Misumi, H. Ohuchi, H. Yoshioka, and S. Noji. 2000. Correlation of diversity of leg morphology in *Gryllus bimaculatus* (cricket) with divergence in *ppp* expression pattern during leg development. *Dev* 127:4373-4381.
- Nulsen, C., and L. M. Nagy. 1999. The role of *wingless* in the development of multibranched crustacean limbs. *Dev Genes Evol* 209:340-348.
- Nusse, R., and H. E. Varmus. 1992. Wnt genes. *Cell* 69:1073-1087.
- Omi, M., M. Fisher, N. J. Maihle, and C. N. Dealy. 2005. Studies on epidermal growth factor receptor signaling in vertebrate limb patterning. *Dev Dyn early view*
- Osorio, D., J. P. Bacon, and P. M. Whittington. 1997. The evolution of arthropod nervous systems. *Am Sci* 85:244-253.
- Padgett, R. W., R. D. St. Johnston, and W. M. Gelbart. 1987. A transcript from a *Drosophila* pattern gene predicts a protein homologous to the transforming growth factor-beta family. *Nature* 325:81-85.
- Padgett, R. W., J. M. Wozney, and W. M. Gelbart. 1993. Human BMP sequences can confer normal dorsal-ventral patterning in the *Drosophila* embryo. *PNAS* 90:2905-2909.
- Page, R. D. M. 2001. TreeViewPPC. University of Glasgow.
- Palopoli, M. F., and N. Patel. 1998. Evolution of the interaction between *Hox* genes and a downstream target. *Curr Biol* 8:587-590.
- Panganiban, G. 2000. *Distal-less* function during *Drosophila* appendage and sense organ development. *Dev Dyn* 218:554-562.
- Panganiban, G., S. M. Irvine, C. Lowe, H. Roehl, L. S. Corley, B. Sherbon, J. K. Grenier, J. F. Fallon, J. Kimble, M. Walker, G. A. Wray, B. J. Swalla, M. Q. Martindale, and S. B. Carroll. 1997. The origin and evolution of animal appendages. *PNAS USA* 94:5162-5166.
- Panganiban, G., L. M. Nagy, and S. B. Carroll. 1994. The role of the *Distal-less* gene in the development and evolution of insect limbs. *Curr. Biol.* 4:671-675.
- Panganiban, G., and J. L. R. Rubenstein. 2002. Developmental functions of the Distal-less/Dlx homeobox genes. *Dev* 129:4371-4386.
- Panganiban, G., A. Sebring, L. Nagy, and S. Carroll. 1995. The development of crustacean limbs and the

- evolution of arthropods. *Science* 270:1363-1366.
- Patel, N. H. 1994a. Pp. 445-487. *Methods in Cell Biology*.
- Patel, N. H. 1994b. *Drosophila melanogaster*: Practical uses in cell biology in J. L. Goldstein and Fryberg, eds. Academic Press.
- Patel, N. H. 1994c. Evolution of insect patterning. *PNAS USA* 91:7385-7386.
- Patel, N. H., E. Martin-Bianco, K. G. Coleman, S. J. Poole, M. C. Ellis, T. B. Kornberg, and C. S. Goodman. 1989. Expression of *engrailed* proteins in arthropods, annelids, and chordates. *Cell* 58:955-968.
- Patrino, M., A. Smertenko, M. D. Candia Carnevali, F. Bonasoro, P. W. Beesley, and M. C. Thorndyke. 2002. Expression of transforming growth factor beta-like molecules in normal and regenerating arms of the crinoid *Antedon mediterranea*: immunocytochemical and biochemical evidence. *Proc R Soc Lond S* 269:1741-1747.
- Pavlopoulos, A., and M. Averof. 2002. Developmental evolution: Hox proteins ring the changes. *Curr. Biol.* 12:R291-R293.
- Pavlopoulos, A., and M. Averof. 2005. Establishing genetic transformation for comparative developmental studies in the crustacean *Parhyale hawaiiensis*. *PNAS* 102:7888-7893.
- Peel, A. 2004. The evolution of arthropod segmentation mechanisms. *BioEssays* 26:1108-1116.
- Peel, A., and M. Akam. 2003. Evolution of segmentation: rolling back the clock. *Current Biology* 13:R708-R710.
- Peel, A., A. D. Chipman, and M. Akam. 2005. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nature Revs Genet*:1-12.
- Peel, A., M. J. Telford, and M. Akam. 2006. The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. *Proc. R. Soc. Lon. B* 273:1733-1742.
- Peifer, M., D. Sweeton, C. M., and E. Wieschaus. 1994. *wingless* signal and Zeste-white 3 kinase trigger opposing changes in the intracellular distribution of Armadillo. *Development* 120:369-380.
- Peifer, M., and E. Wieschaus. 1990. The segment polarity gene *armadillo* encodes a functionally modular protein that is the *Drosophila* homolog of human plakoglobin. *Cell* 63:1167-1178.
- Peltenburg, L. T. C., and C. Murre. 1996. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *The EMBO Journal* 15:3385-3393.
- Penton, A., and F. M. Hoffmann. 1996. Decapentaplegic restricts the domain of *wingless* during *Drosophila* limb patterning. *Nature* 382:162-164.
- Persson, U., H. Izumi, S. Souchelnytskyi, S. Itoh, S. Grimsby, U. Engström, C.-H. Heldin, K. FUna, and P. ten Dijke. 1998. The L45 loop in type I receptors for TGF-beta family members is a critical determinant in specifying Smad isoform activation. *FEBS Letts.* 434:83-87.
- Peterson, M. D., A. Popadic, and T. C. Kaufman. 1998. The expression of two *engrailed*-related genes in an apterygote insect and a phylogenetic analysis of *engrailed*-related genes. *Dev Genes Evol* 208
- Popadic, A., and L. Nagy. 2001. Conservation and variation in *Ubx* expression among chelicerates. *Evol & Dev* 3:391-396.
- Popadic, A., G. Panganiban, D. Rusch, W. A. Shear, and T. C. Kaufman. 1998. Molecular evidence for the gnathobasic derivation of arthropod mandibles and for the appendicular origin of the labrum and other structures. *Dev Genes Evol* 208:142-150.
- Popadic, A., D. Rusch, M. Peterson, B. T. Rogers, and T. C. Kaufman. 1996. Origin of the arthropod mandible. *Nature* 380:395.
- Prpic, N.-M. 2004a. Homologs of *wingless* and *decapentaplegic* display a complex and dynamic expression profile during appendage development in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). *Frontiers in Zoology* 1
- Prpic, N.-M. 2004b. Vergleichende Studien zur Gliedmassenentwicklung bei Arthropoden. Mathematisch-Naturwissenschaftlichen Fakultät. Universität zu Köln, Köln.
- Prpic, N.-M., and W. G. M. Damen. 2004. Expression patterns of leg genes in the mouthparts of the spider *Cupiennius salei* (Chelicerata: Arachnida). *Dev Genes Evol* 214:296-302.
- Prpic, N.-M., R. Janssen, W. G. M. Damen, and D. Tautz. 2005. Evolution of dorsal-ventral axis formation in arthropod appendages: *HL5* and *optomotor-blind/bifid*-type T-box genes in millipede *Glomeris marginata* (Myriapoda: Diplopoda). *Evo & Dev* 7:51-57.
- Prpic, N.-M., R. Janssen, B. Wigand, M. Klinger, and W. G. M. Damen. 2003. Gene expression in spider appendages reveals reversal of *exd/hth* spatial specificity, altered leg gap gene dynamics, and suggests

- divergent distal morphogen signaling. *Dev Biol* 264
- Prpic, N.-M., and D. Tautz. 2003. The expression of the proximodistal axis patterning genes *Distal-less* and *Dachschrund* in the appendages of *Glomeris marginata* (Myriapoda: Diplopoda) suggests a special role of these genes in patterning the head appendages. *Dev Biol* 260:97-112.
- Prpic, N.-M., B. Wigand, W. G. M. Damen, and M. Klinger. 2001. Expression of *dachschrund* in wild-type and *Distal-less* mutant *Tribolium* corroborates serial homologies in insect appendages. *Dev Genes Evol* 211:467-477.
- Prud'homme, B., R. de Rosa, D. Arendt, J.-F. Julien, R. Pajaziti, A. W. Dorresteyn, A. Adoutte, J. Wittbrodt, and G. Balavoine. 2003. Arthropod-like expression patterns of *engrailed* and *wingless* in the annelid *Platynereis dumerilii* suggest a role in segment formation. *Curr. Biol.* 13:1876-1881.
- Prud'homme, B., N. Lartillot, G. Balavoine, A. Adoutte, and M. Vervoort. 2002. Phylogenetic analysis of the Wnt gene family: insights from Lophotrochozoan members. *Curr Biol* 12:1395-1400.
- Quéinnec, E., Mouchel-Vielh, E. et al. 1999. Cloning and expression of the *engrailed.a* gene of the barnacle *Sacculina carcini*. *Dev Genes Evol* 209: 180-185.
- Rattner, A., J.-C. Hsieh, P. M. Smallwood, D. J. Gilbert, N. G. Copeland, N. A. Jenkins, and J. Nathans. 1997. A family of secreted proteins contains homology to the cysteine-rich ligand-binding domain of frizzled receptors. *PNAS* 94:2859-2863.
- Ray, R. P., and K. A. Wharton. 2001. Context-dependent relationships between the BMPs *gbb* and *dpp* during development of the *Drosophila* wing imaginal disk. *Development* 128:3913-3925.
- Raz, E., and B.-Z. Shilo. 1993. Establishment of ventral cell fates in the *Drosophila* embryonic ectoderm requires DER, the EGF receptor homolog. *Genes & Dev* 7:1937-1948.
- Regev, S. 1975. *Environ. Entomol.* 4:307-.
- Regev, S. 1980. *Environ. Entomol.* 9:50-52.
- Reich, A., A. Sapir, and B.-Z. Shilo. 1999. Sprouty is a general inhibitor of receptor tyrosine kinase signaling. *Development* 126:4139-4147.
- Rempel, J. G. 1975. The evolution of the insect head: the endless dispute. *Quaest Entomol* 11:7-25.
- Rida, P. C. G., N. Le Minh, and Y.-J. Jiang. 2004. A Notch feeling of somite segmentation and beyond. *Dev Biol* 265:2-22.
- Riggleman, B., P. Schedl, and E. Wieschaus. 1990. Spatial expression of the *Drosophila* segment polarity gene *armadillo* is posttranscriptionally regulated by *wingless*. *Cell* 63:549-560.
- Rogers, B. T., and T. C. Kaufman. 1996. Structure of the insect head as revealed by the EN protein pattern in developing embryos. *Development* 122:3419-3432.
- Ronquist, F., and J. P. Huesenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Ronquist, F., and J. P. Huesenbeck. 2005. MRBAYES: Bayesian Analysis of Phylogeny. Florida State University, Tallahassee.
- Ronshaugen, M., N. McGinnis, and W. McGinnis. 2002. Hox protein mutation and macroevolution of the insect body plan. *Nature* 415:914-917.
- Röttinger, E., Besnardeau, L. and Lepage, T. 2004. A Raf/MEK/ERK signalling pathway is required for development of the sea urchin embryo micromere lineage through phosphorylation of the transcription factor Ets. *Dev* 131: 1075-1087.
- Rousset, R., J. A. Mack, K. A. Wharton, J. D. Axelrod, K. M. Cadigan, M. P. Fish, R. Nusse, and M. P. Scott. 2001. *naked cuticle* targets *dishevelled* to antagonise Wnt signal transduction. *Genes & Dev* 15:658-671.
- Rutledge, B. J., K. Zhang, E. Bier, Y. N. Jan, and N. Perrimon. 1992. The *Drosophila* spitz gene encodes a putative EGF-like growth factor involved in dorsal-ventral axis formation and neurogenesis. *Genes & Dev* 6:1503-1517.
- Sampath, T. K., K. E. Rashka, J. S. Doctor, and R. F. Tucker. 1993. *Drosophila* Transforming Growth Factor-beta superfamily proteins induce endochronal bone formation in mammals. *PNAS* 90:6004-6008.
- Sanchez-Salazar, J., M. T. Pletcher, R. L. Bennet, S. B. Brown, T. J. Dandamudi, R. E. Denell, and J. S. Doctor. 1996. The *Tribolium decapentaplegic* gene is similar in sequence, structure, and expression to the *Drosophila dpp* gene. *Dev Genes Evol* 206:237-246.
- Sanicola, M., J. J. Sekelsky, S. Elson, and W. M. Gelbart. 1995. Drawing a stripe in *Drosophila* imaginal discs: negative regulation of *decapentaplegic* and *patched* expression by *engrailed*. *Genetics* 139:745-756.
- Scadden, A. D. J., and C. W. J. Smith. 2001. RNAi is antagonised by A->I hyper-editing. *EMBO Reports*

- 2:1107-1111.
- Schejter, E. D. 1986. Alternative 5' exons and tissue-specific expression of the *Drosophila* EGF receptor homolog transcripts. *Cell* 46:1091-1101.
- Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 1999. TreePuzzle. Theoretical Bioinformatics Deutsches Krebsforschungszentrum, Heidelberg.
- Schmitt, S., M. Prestel, and R. Paro. 2005. Intergenic transcription through a Polycomb group response element counteracts silencing. *Genes & Dev* 19:697-708.
- Schöck, F., Purnell, B.A., Wimmer, E.A. and Jäckle, H. 1999. Common and diverged functions of the *Drosophila* gene pair *D-Sp1* and *buttonhead*. *Mech Dev* 89: 125-132.
- Schöck, F., Sauer, F., Jäckle, H. and Purnell, B.A. 1999. *Drosophila* head segmentation factor Buttonhead interacts with the same TATA box-binding protein-associated factors and *in vivo* DNA targets as human Sp1 but executes a different biological program. *PNAS* 96: 5061-5065.
- Scholtz, G. 1997. Cleavage, germ band formation and head segmentation: the ground pattern of the Euarthropoda. Pp. Ch 24 in Fortey and Thomas, eds. *Arthropod relationships*.
- Scholtz, G. 2001. Evolution of developmental patterns in arthropods - the analysis of gene expression and its bearing on morphology and phylogenetics. *Zoology* 103:99-111.
- Scholz, H., E. Sadlowski, A. Klaes, and C. Klämbt. 1997. Control of midline glia development in the embryonic *Drosophila* CNS. *Mech. Dev.* 62:79-91.
- Schoppmeier, M., and W. G. M. Damen. 2001. Double-stranded RNA interference in the spider *Cupiennius salei*: the role of *Distal-less* is evolutionarily conserved in arthropod appendage formation. *Dev Genes Evol* 211:76-82.
- Schoppmeier, M., and W. G. M. Damen. 2005. Expression of Pax group III genes suggests a single-segmental periodicity for opisthosomal segment patterning in the spider *Cupiennius salei*. *Evo & Dev* 7:160-169.
- Schubert, M., L. Z. Holland, M. D. Stokes, and N. D. Holland. 2001. Three amphioxus *Wnt* genes (*AmphiWnt3*, *AmphiWnt5* and *AmphiWnt6*) associated with the tail bud: the evolution of somitogenesis in chordates. *Dev Biol* 240:262-273.
- Schweitzer, R., R. Howes, R. Smith, B.-Z. Shilo, and M. Freeman. 1995. Inhibition of *Drosophila* EGF receptor activation by the secreted protein Argos. *Nature* 376:699-702.
- Schweitzer, R., and B.-Z. Shilo. 1997. A thousand and one roles for the *Drosophila* EGF receptor. *TIG* 13:191-196.
- SDB. 2005. The Interactive Fly.
- Seaver, E. C., and L. M. Kaneshige. 2006. Expression of 'segmentation' genes during larval and juvenile development in the polychaetes *Capitella* sp. I and *H. elegans*. *Dev Biol* 289:179-194.
- Selden, P. A., J. A. Corronca, and M. A. Hünicken. 2005. The true identity of the supposed giant fossil spider *Megarachne*. *Biol. Lett.* 1:44-48.
- Serrano, N., and P. H. O'Farrell. 1997. Limb morphogenesis: connections between patterning and growth. *Curr. Biol.* 7:R186-R195.
- Sha, X., A. M. Brunner, A. F. Purchio, and L. E. Gentry. 1989. Transforming growth factor beta 1: importance of glycosylation and acidic proteases for processing and secretion. *Mol Endocrinol* 3:1090-1098.
- Shandala, T., R. D. Kortschak, S. Gregory, and R. Saint. 1999. The *Drosophila* *dead ringer* gene is required for early embryonic patterning through regulation of *argos* and *buttonhead* expression. *Development* 126:4341-4349.
- Shankland, M., and E. C. Seaver. 2000. Evolution of the bilaterian body plan: what have we learned from annelids? *PNAS* 97:4434-4437.
- Shanmugam, K., M. S. Featherstone, and H. U. Saragovi. 1997. Residues flanking the HOX YPWM motif contribute to cooperative interactions with PBX. *J Biol Chem* 272:19081-19087.
- Shiga, Y., R. Yasumoto, H. Yamagata, and S. Hayashi. 2002. Evolving role of Antennapedia protein in arthropod limb patterning. *Development* 129
- Shilo, B.-Z. 2003. Signaling by the *Drosophila* epidermal growth factor receptor pathway during development. *Exp. Cell Res.* 284:140-149.
- Shilo, B.-Z. 2005. Regulating the dynamics of EGF receptor signaling in space and time. *Development* 132:4017-4027.
- Shilo, B.-Z., and E. Raz. 1991. Developmental control by the *Drosophila* EGF receptor homolog DER. *TIG* 7:388-392.

- Shoemaker, C. B., H. Ramachandran, A. Landa, M. G. dos Reis, and D. Stein. 1992. Alternative splicing of the *Schistosoma mansoni* gene encoding a homologue of epidermal growth factor receptor. *Molecular and Biochemical Parasitology* 53:17-32.
- Shubin, N., C. Tabin, and S. Carroll. 1997. Fossils, genes and the evolution of animal limbs. *Nature* 388:639-648.
- Sidow, A. 1992. Diversification of the *Wnt* gene family on the ancestral lineage of vertebrates. *PNAS* 89:5098-5102.
- Sigma. 1997. DAPI Product Information (Sigma Prod. No. D9542: 4',6-Diamidino-2-phenylindole dihydrochloride hydrate). Sigma.
- Simcox, A. A., I. J. H. Roberts, E. Hersperger, M. C. Gribbin, A. Shearn, and J. R. S. Whittle. 1989. Imaginal discs can be recovered from cultured embryos mutant for the segment-polarity genes *engrailed*, *naked* and *patched* but not from *wingless*. *Development* 107:715-722.
- Simonnet, F. 2005. L'approche évo-dévo des chélicérates: étude de scorpion *Euscorpius flavicaudis*. Pp. 158. Evolution et Développement. L'Université Pierre et Marie Curie, Paris.
- Simonnet, F., M.-L. Célrier, and E. Quéinnec. 2005a. *orthodenticle* and *empty spiracles* expression pattern in a spider and a scorpion: implications for chelicerate evolution. in press
- Simonnet, F., J. S. Deutsch, and E. Quéinnec. 2004. *hedgehog* is a segment polarity gene in a crustacean and a chelicerate. *Dev Genes Evol* 214:537-545.
- Simonnet, F., N. Rabet, and E. Quéinnec. 2005b. *caudal* expression during scorpion development. in press
- Skeath, J. B. 1998. The *Drosophila* EGF receptor controls the formation and specification of neuroblasts along the dorsal-ventral axis of the *Drosophila* embryo. *Development* 125:3301-3312.
- Spencer, F. A., F. M. Hoffmann, and W. M. Gelbart. 1982. *Decapentaplegic*: A gene complex affecting morphogenesis in *Drosophila melanogaster*. *Cell* 28:451-461.
- Sprecher, S. G., and H. Reichert. 2003. The urbilaterian brain: developmental insights into the evolutionary origin of the brain in insects and vertebrates. *Arth. Str. & Dev.* 32:141-156.
- St Johnston, D., and W. M. Gelbart. 1987. *Decapentaplegic* transcripts are localised along the dorsal-ventral axis of the *Drosophila* embryo. *EMBO J* 6:2785-2791.
- Stemerdink, C., and J. R. Jacobs. 1997. Argos and Spitz group genes function to regulate midline glial cell number in *Drosophila* embryos. *Development* 124:3787-3798.
- Sternberg, P. W., and H. R. Horvitz. 1991. Signal transduction during *C. elegans* vulval induction. *TIG* 7:366-371.
- Stollewark, A., M. Schoppmeier, and W. G. M. Damen. 2003a. Involvement of *Notch* and *Delta* genes in spider segmentation. *Nature* 423:863-865.
- Stollewark, A., D. Tautz, and M. Weller. 2003b. Neurogenesis in the spider: new insights from comparative analysis of morphological processes and gene expression patterns. *Arth. Str. & Dev.* 32:5-16.
- Stopper, G. F., and G. P. Wagner. 2005. Of chicken wings and frog legs: A smorgasbord of evolutionary variation in mechanisms of tetrapod limb development. *Dev Biol* 288:21-39.
- Suga, H., K. Ono, and T. Miyata. 1999. Multiple TGF-beta receptor related genes in sponge and ancient gene duplications before the parazoan-eumetazoan split. *FEBS Letts.* 453:346-350.
- Sugiura, R., and e. al. 2003. Feedback regulation of MAPK signalling by an RNA-binding protein. *Nature* 424:961-965.
- Suzuki, Y., and M. F. Palopoli. 2001. Evolution of insect abdominal appendages: are prolegs homologous or convergent traits? *Dev Genes Evol* 211:486-492.
- Swofford, D. L. 1998. PAUP: Phylogenetic Analysis Using Parsimony. Sinauer Associates, Sunderland, Massachusetts.
- Tabata, T., and T. Y. 2004. Morphogens, their identification and regulation. *Development* 131:703-712.
- Takeichi, M. 1991. Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* 251:1451-1455.
- Tautz, D. 2004. Segmentation. *Developmental Cell* 7:301-312.
- Telford, M. J., and R. H. Thomas. 1995. Demise of the Atelocerata? *Nature* 376:123-124.
- Telford, M. J., and R. H. Thomas. 1998a. Expression of homeobox genes show chelicerate arthropods retain their deuterocerebral segment. *PNAS USA* 95:10671-10675.
- Telford, M. J., and R. H. Thomas. 1998b. Of mites and *zen*: expression studies in a chelicerate arthropod confirm *zen* is a divergent Hox gene. *Dev Genes Evol* 208:591-594.
- Thomas, R. H., and M. J. Telford. 1999. Appendage development in embryos of the oribatid mite *Archegozetes longesetosus* (Acari, Oribatei, Trhypochthoniidae). *Acta Zoologica* 80:193-200.

- Thorpe, C. J., G. Weidinger, and R. T. Moon. 2005. Wnt/beta-catenin regulation of the Sp1-related transcription factor *sp5l* promotes tail development in zebrafish. *Development* 132:1763-1772.
- Tickle, C. 2003. Patterning systems - from one end of the limb to the other. *Developmental Cell* 4:449-458.
- Tickle, C., and A. Münsterberg. 2001. Vertebrate limb development - the early stages in chick and mouse. *Curr. Opin. Genes & Dev.* 11:476-481.
- Tolkunova, E. N., M. Fujioka, M. Kobayashi, D. Deka, and J. B. Jaynes. 1998. Two distinct types of repression domain in Engrailed: one interacts with the Groucho corepressor and is preferentially active on integrated target genes. *Mol Cell Biol*
- Tolwinski, N. S., and E. Wieschaus. 2004. Rethinking Wnt signaling. *TIG* 20:177-181.
- Tomoyasu, Y., S. R. Wheeler, and R. E. Denell. 2005. *Ultrabithorax* is required for membranous wing identity in the beetle *Tribolium castaneum*. *Nature* 433:643-647.
- Trieckel, D., F. Schöck, H. Jäckle, P. Gruss, and A. Mansouri. 2003. *mBtd* is required to maintain signaling during murine limb development. *Genes & Dev* 17:2630-2635.
- Tsruya, R., Schlesinger, A. et al. 2002. Intracellular trafficking by Star regulates cleavage of the *Drosophila* EGF receptor ligand Spitz. *Genes & Dev* 16: 222-234.
- Udolph, G., J. Urban, G. Rüsling, K. Lüer, and G. M. Technau. 1998. Differential effects of EGF receptor signaling on neuroblast lineages along the dorsoventral axis of the *Drosophila* CNS. *Development* 125:3291-3300.
- Ueno, K., C.-C. Hui, M. Fukuta, and Y. Suzuki. 1992. Molecular analysis of the deletion mutants in the E homeotic complex of the silkworm *Bombyx mori*. *Development* 114:555-563.
- Urban, S., G. Brown, and M. Freeman. 2004. EGF receptor signaling protects smooth-cuticle cells from apoptosis during *Drosophila* ventral epidermis development. *Development* 131:1835-1845.
- Vachon, G., B. Cohen, C. Pfeifle, M. E. McGuffin, J. Botas, and S. M. Cohen. 1992. Homeotic genes of the Bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. *Cell* 71:437-450.
- Vervoort, M. 2002. Functional evolution of Hox proteins in arthropods. *BioEssays* 24:775-779.
- von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research* 14:4683-4690.
- Walker, R. A. 1998. The erbB/HER type 1 tyrosine kinase receptor family. *Journal of Pathology* 185:234-235.
- Walldorf, U., P. Binner, and R. Fleig. 2000. Hox genes in the honey bee *Apis mellifera*. *Dev Genes Evol* 210:483-492.
- Wallingford, J. B., and R. Habas. 2005. The developmental biology of Dishevelled: an enigmatic protein governing cell fate and cell polarity. *Development* 132:4421-4436.
- Waloszek, D., and J. A. Dunlop. 2002. A larval sea spider (Arthropoda: Pycnogonida) from the Upper Cambrian 'Orsten' of Sweden, and the phylogenetic position of pycnogonids. *Palaeontology* 45:421-446.
- Wang, Y.-C., and E. L. Ferguson. 2005. Spatial bistability of Dpp-receptor interactions during *Drosophila* dorsal-ventral patterning. *Nature* 434:229-234.
- Wanninger, A., and G. Haszprunar. 2001. The expression of an *engrailed* protein during embryonic shell formation in the tusk-shell, *Antalis entalis* (Mollusca, Scaphapoda). *Evo & Dev* 3:312-321.
- Weatherbee, S. C., and S. B. Carroll. 1999. Selector genes and limb identity in arthropods and vertebrates. *Cell* 97:283-286.
- Wedeer, C. J., R. G. Kostriken, D. Leach, and P. Whittington. 1997. Segmentally reiterated expression of an engrailed-class gene in the embryo of an Australian onychophoran. *Dev Genes Evol* 270:282-286.
- Wedeer, C. J., and D. A. Weisblat. 1991. Segmental expression of an *engrailed*-class gene during early development and neurogenesis in an annelid. *Development* 113:805-814.
- Weidinger, G., C. J. Thorpe, K. Wuennenberg-Stapleton, J. Ngai, and R. T. Moon. 2005. The Sp1-related transcription factors *sp5* and *sp5-like* act downstream of Wnt/beta-catenin signaling in mesoderm and neuroectoderm patterning. *Curr. Biol.* 15:489-500.
- Weiss, K.M. and Fullerton, S.M. 2000. Phenogenetic drift and the evolution of genotype-phenotype relationships. *Theor. Pop. Biol.* 57: 187-195.
- Weitzel, H. E., M. R. Illies, C. A. Byrum, R. Xu, A. H. Wikramanayake, and C. A. Ettensohn. 2004. Differential stability of beta-catenin along the animal-vegetal axis of the sea urchin embryo mediated by dishevelled. *Development* 131:2947-2956.
- Wennerberg, K., and e. al. 2003. Rnd proteins function as RhoA antagonists by activating p190 RhoGAP.

- Curr Biol 13:1106-1115.
- Wheeler, W. C., G. Giribet, and G. D. Edgecombe. 2004. Arthropod Systematics. Pp. 281-295 in J. Cracraft and M. J. Donoghue, eds. Assembling the Tree of Life. Oxford University Press.
- Wheeler, W. C., and C. Y. Hayashi. 1998. The phylogeny of the extant chelicerate orders. Cladistics 14:173-192.
- Whittington, P. M., T. Meier, and P. King. 1991. Segmentation, neurogenesis and formation of early axonal pathways in the centipede, *Ethmostigmus rubripes*. Roux's Arch Dev Biol 199:340-363.
- Whitman, M., and L. A. Raftery. 2005. TGF-beta signaling at the summit. Development 132:4205-4210.
- Wikramanayake, A. H., M. Hong, P. N. Lee, M. Martindale, and e. al. 2003. An ancient role for nuclear beta-catenin in the evolution of axial polarity and germ layer segregation. Nature 426:446-450.
- Wilkins, A. S. 2002. The Evolution of Developmental Pathways. Sinauer Associates Inc.
- Williams, T. A., and L. M. Nagy. 2001. Developmental modularity and the evolutionary diversification of arthropod limbs. J Exptal Zoo (Mol Dev Evol) 291:241-257.
- Wilson, K. A., M. E. Andrews, F. R. Turner, and R. A. Raff. 2005. Major regulatory factors in the evolution of development: the roles of *gooseoid* and *Max* in the evolution of the direct-developing sea urchin *Heliocidaris erythrogramma*. Evo & Dev 7:416-428.
- Wimmer, E. A. 2005. Crustacean and beetle Sp genes, pers. com.
- Wimmer, E. A., S. M. Cohen, H. Jäckle, and C. Desplan. 1997. *buttonhead* does not contribute to a combinatorial code proposed for *Drosophila* head development. Development 124:1509-1517.
- Wimmer, E. A., G. Frommer, B. A. Purnell, and H. Jäckle. 1996. *buttonhead* and *D-Sp1*: a novel *Drosophila* gene pair. Mechanisms of Development 59:53-62.
- Wimmer, E. A., M. Simpson-Brose, S. M. Cohen, C. Desplan, and H. Jäckle. 1995. *Trans*- and *cis*-acting requirements for blastodermal expression of the head gap gene *buttonhead*. Mechanisms of Development 53:235-245.
- Wodarz, A., and R. Nusse. 1998. Mechanisms of Wnt signaling in development. Ann. Rev. Cell. Dev. Biol. 14:59-88.
- Wray, C. G. 2007. The evolutionary significance of *cis*-regulatory mutations. Nature Reviews Genetics 8:206-216.
- Wray, C. G., D. K. Jacobs, R. Kostriken, A. P. Vogler, R. Baker, and R. DeSalle. 1995. Homologues of the *engrailed* gene from five molluscan classes. FEBS Letts. 365:71-74.
- Xie, Z., K. D. Kasschau, and J. C. Carrington. 2003. Negative feedback regulation of *Dicer-Like1* in *Arabidopsis* by microRNA-guided mRNA degradation. Curr Biol 13:784-789.
- Yabe, T., T. Shimizu, and e. al. 2003. Ogon/Secreted frizzled functions as a negative feedback regulator of Bmp signaling. Development 130:2705-2716.
- Yagi, Y., T. Suzuki, and S. Hayashi. 1998. Interaction between *Drosophila* EGF receptor and *vnd* determines three dorsoventral domains of the neuroectoderm. Development 125:3625-3633.
- Yamamoto, D. S., M. Sumitani, K. Tojo, J. M. Lee, and M. Hatakeyama. 2004. Cloning of a *decapentaplegic* orthologue from the sawfly, *Athalia rosae* (Hymenoptera), and its expression in the embryonic appendages. Dev Genes Evol 214:128-133.
- Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell. 1984. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell 39:27-38.
- Yamamoto, T., T. Nishida, N. Miyajima, S. Kawai, T. Ooi, and K. Toyoshima. 1983. The *erbB* gene of avian erythroblastosis virus is a member of the *src* gene family. Cell 35:71-78.
- Yang, J., J. Wu, C. Tan, and P. S. Klein. 2003. PP2A:B56-epsilon is required for Wnt/beta-catenin signaling during embryonic development. Dev 130:5569-5578.
- Younossi-Hartenstein, A., P. Green, G.-J. Liaw, K. Rudolph, J. Lengyel, and V. Hartenstein. 1997. Control of early neurogenesis of the *Drosophila* brain by the head gap genes *tll*, *otd*, *ems* and *bt2*. Dev Biol 182:270-283.
- Zeng, W., K. A. Wharton Jr, J. A. Mack, K. Wang, M. Gadbaw, K. Suyama, P. S. Klein, and M. P. Scott. 2000. *naked cuticle* encodes an inducible antagonist of Wnt signalling. Nature 403:789-795.
- Zeng, Y. A., and E. M. Verheyen. 2004. Nemo is an inducible antagonist of Wingless signaling during *Drosophila* wing development. Development 131:2911-292-.
- Zhang, H., Y. Shinmyo, T. Mito, K. Miyawaki, I. Sarashina, H. Ohuchi, and S. Noji. 2005. Expression patterns of the homeotic genes *Scr*, *Antp*, *Ubx* and *abd-A* during embryogenesis of the cricket *Gryllus*

- bimaculatus*. Gene Expression Patterns 5:491-502.
- Zhang, Z.-Q., and R. J. Jacobson. 2000. Using adult female morphological characters for differentiating *Tetranychus urticae* complex (Acari: Tetranychidae) from green-house tomato crops in the UK. Systematic & Applied Acarology 5:69-76.
- Zheng, Z., A. Khoo, D. Fambrough, L. Garza, and R. Booker. 1999. Homeotic gene expression in the wild-type and a homeotic mutant of the moth *Manduca sexta*. Dev Genes Evol 209:460-472.

APPENDICES



Faucheur commun

Appendix 1: Spider mite Chaetotaxy²⁷

What is chaetotaxy?

Various types of sensory hairs or structures are present on the dorsum and walking appendages of *Tetranychus urticae* adults. Chaetotaxy refers to the number, arrangement and form of sensory setae, the various types of which are made of actinochitin. Actinochitin is birefringent in polarised light, and unique to the Acariformes such that they are also termed 'Actinotrichida', the less derived suborders Opilioacariformes and Parasitiformes being classed 'Anactinotrichida' (Brusca and Brusca 2003; Brussels 2006). Distinct types of sensory seta are dedicated to olfactory, gustatory, tactile and mechanosensory perception, with dorsal limb fissures and setal clusters being employed to detect strain, vibration and air flow (Barth and Stagl 1976; Brusca and Brusca 2003; Grandjean 1935; Grandjean 1936; Grandjean 1939).

Chaetotaxy in *Tetranychus*

Chaetotactic features specific to *Tetranychus urticae* include empodium morphology and a proximal 'twin cluster' (tactile seta plus solenidium) on the L1 tarsus, distal to tactile setae adjacent to the tibia (Flechtmann and Knihinicki 2002; Grandjean 1948; Zhang and Jacobson 2000). Setae are opaque, hair-like, birefringent projections with a plugged root, whereas solenidia are shorter, microporous, monorefringent projections with a hollow root and usually function as olfactory sensors but when combine in a duplex/twin cluster may be sensitive to vibrations or air flow (Brussels 2006). Empodia are present at the distal tip of legs L1-L4, composed of 3 pairs of setae with a very short, uncurved spur that defines the species - as empodial spurs are normally curved and extended to at least one-third the length of the setal hairs (c.f Figure 2.2.4b, iv). Furthermore regarding chaetotaxy, a single pair of para-anal setae is diagnostic at the genus level and *T. urticae* dimorphs are distinguishable as the L1 tibia of females has fewer pairs of setae than the males (Flechtmann and Knihinicki 2002; Zhang and Jacobson 2000).

²⁷ c.f. Chapter II, section 2.2.4

Appendix 2: Oligonucleotide Primers

2.1 *Distal-less* primers

Dll degenerate PCR primers

degDll F1	5' -CGT AAR CCN MGN CAN ATH TA-3'
degDll F2	5' -CGT AAG CCG MGN CAN ATH TA-3'
degDll R1	5' -ACG RTT YTG RAA CCA DAT YTT-3'
CCF Dll F1	5' -CAG TAT CTG GCA YTN CCN GAR-3'
CCF Dll F2	5' -GGT TCC AAC GNA CNC ART AY-3'
CCF Dll F3	5' -AAT GGC AAG GGN AAR AAR ATG-3'
CCF Dll R1	5' -CCA TAT CTT NAC YTG NGT YTG-3'
CCF Dll R2	5' -TTG GAC CGT CKR TTY TGR AAC CA-3'
CCF Dll R3	5' -GCT TTC ARC ATY TTY TTR TA-3'

Tu-Dll-specific primers for cDNA library screening probe amplification

Tudll U39	5' TTC AAC AGC TTA ACA GAC G 3'
Tudll L116	5' CCT GTG TTT GAG TCA ATC C 3'

Tu-Dll-specific primers for in situ probe DNA amplification

Dll U41	5' CCA AGA GAC GAG AAA CCA AC 3'
Dll L845	5' TGG TGA TTC AGA GTA AAC CG 3'
Dll U101	5' AAC TGG AAA CCT CGT ACA AT 3'
Dll L790	5' GCT GTT GAG GAA GAG GAA GA 3'
Dll U297	5' TGA ATG GAG GTG GAC AAG TG 3'
Dll L608	5' GGG GAA TGT TGA GGA ATG TA 3'

2.2 *Sp* primers

Degenerate primers for Sp gene screening

Dmbtd-R1	5' A TGT GTY TTY TTR TGY TT 3'
btdSp-R2	5' A TGC GTY TTN ACR TGY TT 3'
btdSp-R3	5' AAG TGG TCA WSN GKC ATR AA 3'
btdSp-F1	5' TTG AAA GCA CAY YTN MGN TGG 3'
btdSp-F2	5' TGC CAC ATH CCN GGN TGY GG 3'
btdSp-F3	5' TGC GAC TGY CCN AAY TGY MG 3'

Sp-F1 ²⁸	5' GGN AAR GGN TTY CAY CCN TGG AA 3'
Sp-F1b	5' GCN CAN TGY GAY TGY CCN AAY TG 3'
Sp-F2	5' TGY CCN AAY TGY CAR GAR GCN GA 3'
Sp-R1	5' GGR CAN GCR AAN CKY TTY TCN CC 3'
Sp-R2	5' AAN GGN CKY TCN CCN GTR TGC CA 3'

²⁸ Sequences for primers Sp-F1, Sp-F1b, Sp-F2, Sp-R1, Sp-R2 courtesy of Martin Hildebrant, Universität zu Köln.

2.3 *Wingless/Wnt1* primers

Degenerate primers for Wg/Wnt1 gene screening

CCF Wg F1	5'-TGT ACC GTA AAR CAN TGY TGG ATG-3'
CCF Wg R1	5'-GTC GAC CCC TAT GSW NGT RTC RTT-3'
CCF Wg R2	5'-TAG GTC GCA CCC RTC NAC NCC-3'
CCF Wg R3	5'-GTT CGG TAC CCT CGN CCR CAR CAC AT-3'
Cs Wg F1 ²⁹	5'-ATH GAR WSN TGY CAN GRY GAY TA-3'
Cs Wg F2	5'-TGG GAR TGG GGN GGN TGY WSN GA-3'
Cs Wg R1	5'-ACY TWR CAR CAC CAN TGR AAN GTR CA-3'
Cs Wg R2	5'-ACY TWR CAR CAC CAR TGR AAN GTR CA-3'
ADP Wg F1 ³⁰	5'-GGG CGA CAA CCT GAA GGA CHS NTT YGA YGG-3'
ADP Wg F2	5'-GTG CAC GAT CCG CAC GTG CTG GAT GMG NHT-3'
ADP Wg R1	5'-CGA CGC CGA TGG AGG TNT CRT TRC-3'
ADP Wg R2	5'-GCC CGC AGC ACA TGA GGY CRC ANC CRT-3'
ADP Wg R3	5'-TGC GGT AGC CGC GCS HRC ARY ACA T-3'

2.4 *Engrailed* primers

Degenerate primers for En screening

En EH1-Fwd ³¹	5' ATA TTR AGR CCN GAR TTY GG 3'
En EH2-Fwd	5' GG CCC GCC TGG GTN TAY TG 3'
En EH5-Rev	5' TGT GAG TGG TTG TAC AGG CCY TGN GCC 3'

Tu-En1 and Tu-En2 gene-specific iPCR primers

iEn-KD U241	5' CAA GCG AGC CAA AAT AAA G 3'
iEn-KD L40	5' AGA TCG AGG ACC TGA AGA C 3'
iEn-KD U208	5' GAA CGA GTC GCA AAT CAA A 3'
iEn-KD L108	5' AGT TGC TCT GCG GTG AAT G 3'
iEn-KS U240	5' CAG AAT AAG CGG GCC AAG AT 3'
iEn-KS L21	5' GGT CTA TCC GAG TAT CCG AGT 3'
iEn-KS U173	5' CGA GAA ACG TCG TCA AGA T 3'
iEn-KS L131	5' CTC TTG TTT CAA TCG TTG C 3'

Tu-En1 gene-specific primers for in situ probe DNA amplification

EnTu U56	5' GCT AAT TTC AGT TTT TGC TCA 3'
EnTu L948	5' AAC TTT TCT CGC ATC AAA TTA 3'
EnTu U225	5' ACT TTT GAC TTT TGC TGT TTT 3'
EnTu L1041	5' GTT CAA CTT TAA ATC TCG AGC 3'
EnTu U400	5' CCA CTA ACC CAA TCA AAA GT 3'
EnTu L825	5' TCC TTT TTT CTT AGC TTT CTC 3'
EnTu U311	5' CTC AAC CTT ACC TTA CCT TTT 3'
EnTu L1165	5' TCA TTT GTT GGA GTT TTC ATC 3'
EnTu U972	5' TCA TTT GTT GGA GTT TTC ATC 3'

²⁹ Sequences for Cs-Wg primers from Damen (2002).

³⁰ Sequences for ADP-Wg primers courtesy of Andrew Peel, University of Cambridge.

³¹ Sequences for En-EH primers courtesy of A. Peel.

EnTu L375	5' GA ATA CCG ATG AAA CCA TGT 3'
EnTu U448	5' CAC CTG AAT TTT GTA AAT CGA 3'
EnTu L972	5' GAT GAA AAC TCC AAC AAA TGA 3'

2.5 Decapentaplegic primers

Dpp degenerate PCR screen primers

CCF Dpp F1	5'-CCA CTG TTT ACC TAY CAN GAY GAY GG-3'
CCF Dpp F2	5'-TGG GAT GAC TGG ATH GTN GCN CC-3'
CCF Dpp R1	5'-TCG CAG ACT TGC TGN GCR TGR TTN GT-3'
CCF Dpp R2	5'-CAT CTC TTG RTA RTT YTT NAR-3'
CCF Dpp R3	5'-CCG GCA TCC GCA NCC NAC NAC-3'
degdpp R3	5'-CAN CAY CCN ACR CCC CAA ACG GTC ATC TCC T-3'
degdpp R2	5'-CCC ACC ACG GTC ATC TCC TKR TAR TTY TT-3'
degdpp F2	5'-GGC GGC ACC CCA TGT AYG TNG AYT T-3'
degdpp R1	5'-GGG TCT GCA CCA CGG CRT GRT TNG T-3'
degdpp F1	5'-AGC CCC TGC TGT TCA CCT AYA CNG AYG A-3'

Tu-dpp gene-specific iPCR primers

iDpp U111	5' TTC CAA CCA ACT TCA CTA 3'
iDpp L33	5' TCT CCT CAA CGA TGT CCA 3'
iDpp L56	5' TAT TGT GCG GGT GAC TGT 3'
iDpp U61	5' CAC CCG CAC AAT AAA ATG 3'
iDpp U77	5' AGA TAC ATC ATG CCC CGA 3'
iDpp L47	5' CAT GGT GCA CAG AGA TCT 3'

Tu-dpp-specific primers for library screening probe amplification

Tudpp U26	5' GGC ATT ATG GAC ATC GTT 3'
Tudpp L111	5' TTA GTG AAG TTG GTT GGA A 3'

Tu-dpp-specific primers for in situ probe DNA amplification

Tu-dpp U11	5' GGC ACG AGG TTT ACT AAC AC 3'
Tu-dpp L1028	5' TTT ATT TGG TTT ACC TTT TG 3'
Tu-dpp U87	5' TTC TAC ATT GAG ATT GCG GT 3'
Tu-dpp L959	5' TGT GTT TGG TTT GGT TGA TA 3'
Tu-dpp U437	5' TTT TAG TGA AGT TGG TTG GA 3'
Tu-dpp L745	5' GCT GAT GTT TTG TTG TGA AT 3'
Tu-dpp U194	5' GAG TCT GGC TCT TGA CGA T 3'
Tu-dpp L854	5' CAA AAT GGC AAA CAA AAG G 3'

2.6 Epidermal Growth Factor Receptor primers

i) *TER-a*

Degenerate EGFR screening primers that amplified TER-a

degEGFR F1	5'-AAG AAC GTG AAG ATC CCC GTN GCN ATH AA-3'
degEGFR R2	5'-TC CGG GCG GCC ARR TCN CKR TG-3'

TER-a gene-specific iPCR primers

iTER1 L43	5' TTG TAT CTG GAC AGT GCC 3'
iTER1 L78	5' TGC CAT GAT GTA AGC CTC 3'
iTER1 L166	5' CAC CCT AAA GGC ATA AGT T 3'
iTER1 U219	5' ATT GGC TCA AAA CCT TTA T 3'
iTER1 U256	5' TTG CTA GAG GAA TGG CTT 3'
iTER1 U277	5' TTG AAG ATA AAC GAA TGG T 3'

TER-a specific primers for in situ probe DNA amplification

TER1 U95	5' TGC ACT GCT TAT GGT ATT 3'
TER1 U38	5' TTG AAA TCG TTG TCT GGT 3'
TER1 U504	5' TTC AGT TAC CAG GGT CTT 3'
TER1 U767	5' GGC ACT GCT CCA GAT ACA 3'
TER1 L1407	5' CAA ACA CCA TAG CCA AGT 3'
TER1 L1450	5' CCT GGA ATG ACG AGA TAT 3'
TER1 L987	5' CAA GAT AAG CCA TTC CTC 3'
TER1 L772	5' TTG TTT GTA TCT GGA GCA 3'

ii) 'TER-b'***Degenerate EGFR screening primers that amplified putative TER-2***

degEGFR F2	5'-AAC GTG AAG ATC CCN GTN GCN AT-3'
degEGFR R1	5'-CGT GCT TGG GTG TCT GCA CAN RNA CRT T-3'
degEGFR R3	5'-GGT CAC GCC GAA GSH CCA NAC RTC-3'

TER-b gene-specific iPCR primers

iTER-2 U289A	5' TAT TCG GCT GCG TCC CTT 3'
iTER-2 L143	5' CTC TTG GAC TAC GTG CGA 3'
iTER-2 U160	5' GAC AAC CCA GCG GCA TCA 3'
iTER-2 L101	5' CGA AAC CTC TAC TCA ACT GG 3'
iTER-2 U289'	5' TAT TCG GCT GCG TCC CTT C 3'
iTER-2 U189	5' AGC ATC AGT TGG GAG GTC AT 3'
iTER-2 L119	5' AAA GAC AAA ATC GGT TCG 3'
iTER-2 L97	5' CCT CTA CTC AAC TGG TGC A 3'
iTER-2 U152	5' GCT GGG TTG TCT CTT GGA C 3'

TER-b specific cDNA library screening probe amplification primers

TER2 U285	5' ATT CCT TGT TCG TAT TCG G 3'
TER2 U60	5' CCA TGC GTT TCT CCT CTA G 3'

TER-b specific primers to re-test for presence in Tetranymphus genome

TER2 U8	5' AGC CGA ATA CGA ACA AGG A 3'
TER2 L218	5' AGA TAC ATC ATG CCC CGA G 3'
TER2 U34	5' GAG GAA GCC TAC ATC ATG G 3'
TER2 L236	5' ACC ATG CGT TTC TCC TCT A 3'
TER2 U92	5' TCT GCA TGA CCT CCC AAC T 3'
TER2 L158	5' TTG TCT TTG TTG TTT CGC A 3'

2.7 Hox primers

Degenerate Hox screen primers targetted against homeobox sequences

LELE/An1 Hox F1 ³²	5'-CG GAT TCC CTA GAG CTN GAR AAR GAR T-3'
QNKR/An2 Hox R1	5'-GGA ATT CAT ICK ICK RTT YTG RAA CCA IAT YT-3'
98C Hox F2 ³³	5'-G GAA TTC GAR CTI GAR AAR GAR TT-3'
97I Hox R2	5'-G CTC TAG ACG ICG RTT TTG RAA CCA-3'

Degenerate Hox primers against YPWM motif e² linker sequences

YPWM F1	CCA ATC CCG GAY GTN CCN TA (PIPDVPY)
YPWM F2	CCC GAT GTA CCN TAY CCN TGG ATG (PDVPYPWM)
YPWM F3	ATC TCT ATA GCG GGN CCN AAY GG (MSIAGPNG)
YPWM F4	CCG AAC GGT TGY CCN MGN MG (PNGCPRR)
YPWM F5	GA GGC AGR CAR CAN TAY AC (RGRQTYT)
YPWM F6	ACG CGC TTC CAR CAN YTN GA (TRFQTL E)

Tu-Ubx iPCR primers

iUbx	iUbx L60	5' CAA AGA GCG TGA GCC AGT 3'
	iUbx U58	5' GAA CTG GCT CAC GCT CTT 3'
	iUbx U85	5' GAA CGT CAG ATT AAA ATC 3'
	iUbx U73	5' CTT TGC CTA AGT GAA CGT 3'
	iUbx L14	5' GAC TTG TGT GAA ACT CCT 3'
	iUbx L51	5' TGA GCC AGT TCA ATG CGA 3'

Tu-Scr iPCR primers

iDfd-Scr L56	5' CAA TTT CAA TAC GCC GAC 3'
iDfd-Scr L39	5' CGA CGG GTT AAA TAA CGA 3'
iDfd-Scr L52	5' TTC AAT ACG CCG ACG ACG 3'
iDfd-Scr U88	5' CTA GGA AAG ACA GAT CAA 3'
iDfd-Scr U72	5' TGC TCA TTC ACT TTG CCT 3'
iDfd-Scr U103	5' CAA RAT CTG GTT CCA AAA 3'

Tu-Ubx specific primers for in situ probe DNA amplification

TuUbx F25	5' TCA AAA AAA GCC ATT AGC T 3'
TuUbx F61	5' GCG TCC AAA GAG TCA AAG T 3'
TuUbx F162	5' CCT TTC TCT TTC ACT CCT T 3'
TuUbx F201	5' CTC TCT CTT TCA ATG GTG C 3'
TuUbx R712	5' GGA TTC TTT TGT CGC TAT T 3'
TuUbx R762	5' CGA TGG ATG ATG ATT TGT G 3'
TuUbx R780	5' ATG ATG GTG ATG AAT GGC 3'
TuUbx R823	5' TTT GTG TGT CAG TGT TGG C 3'

³² Sequences for primers Hox-F1 and Hox-R1 from Martinez et al. (1997).

³³ Sequences for primers Hox-F2 and Hox-R2 from Cook et al. (2001).

Appendix 3: Non-target gene sequences and analysis

3.1 *Tetranychus* receptor tyrosine kinase (RTK) *Src*³⁴

During degenerate PCR screening for EGFR (section 5.4.1) one non-target gene was also identified; a putative *Tetranychus Src* ortholog (Figure 3.1a). *Src* codes for a tyrosine kinase named after Schmidt-Ruppin viral oncogene/avian sarcoma virus. The *Src* gene family comprises several related oncogenes, those which display tyrosine kinase activity including *EGFR*, *Src* and *abelson* (Livneh et al. 1985; Yamamoto et al. 1983). I amplified *Tu-Src* gene with the distinct primer combination EGFR-F2 and EGFR-R2, presumably degeneracy (°D=64 and °D=32 respectively) allowing recognition of non-target but closely related *EGFR*-like RTK sequence motifs. Although its affinity was speculative from the outset, being approximately 100bp larger than any predicted target *EGFR* fragment, I cloned the 400-450bp PCR product to compare with other products and to gauge the nature of any non-specific binding attributable to degenerate *EGFR* priming.

Tu-Src sequence analysis

Figure 3.1b illustrates nucleotides (429bp) and associated translation (143 amino acids) for the putative *Tu-Src* gene, including a protein sequence alignment comparing 90 homologous amino acids from 4 domains within the *EGFR* and *Src* genes of *Tetranychus urticae* and *Drosophila melanogaster*. Whereas *TER-a* and *TER-b* share 88.8% similarity at the amino acid level (80 of 90 compared), *TER-a* and *Tu-Src* share only 50% amino acid identity (45/90), and *TER-b* and *Tu-Src* only 45.5% (41/90). Between *Tetranychus* and *Drosophila* orthologs of both *EGFR* and *Src*, 40% (i.e. 36/90) of residues are conserved (see Figure 3.1b). These values support the idea that although *EGFR* and *Src* are related within the same oncogene family (and hence share amino acids within the intracellular TK domain as stabilising selection acts to preserve the kinase structure, and critical mitogenic, function), molecular evolution has permitted substantial divergence (38-43%) within each gene lineage (Yamamoto et al. 1983).

I constructed a larger multiple protein sequence alignment (c.f Figure 5.4.1c) comparing 90 amino acids from 4 consecutive *Src* family orthology domains within the intracellular tyrosine kinase domain, including a range of *EGFR* orthologs and *EGFR*-related tyrosine kinase *Src* and *Abelson* groups (Livneh et al. 1985). Within the *EGFR* orthology group, 30% (27/90) of residues are identical, whereas shared identity is reduced to 22.2% (20/90) when *EGFR*, *Src* and *Abelson*

³⁴ c.f. Chapter V section 5.4.1

(i) *Tetranychus Src*

1/1
AAG AAC GTG AAG ATC CCC GTG GCG ATC AAA ACA CTC AAA CCA GGC ACA ATG GAT CCG AAA GAC TTC TTG GAA GAA GCC CAA ACC ATG AAG
lys asn val lys ile pro val ala ile lys thr leu lys pro gly thr met asp pro lys asp phe leu glu glu ala gln thr met lys
91/31
TAC GCA GTG TGC ACC CTC GAA GAA CCA ATC TAC ATT ATC ACG GAG CTC ATG AAG AAC GGA AAA TTA TCG AGA CAT CCG AAA CTG GTT CAG CTG
tyr ala val cys thr leu glu glu pro ile tyr ile ile thr glu leu met lys asn gly lys leu arg his pro lys leu val gln leu
181/61
TCT CTG CTT GAG TAC TTG CAA GGT AGG GCT CTT CAT CTA GGT CCT AAC ACT GCC CTA GCA TTA TCG CTA GGA ATG AAA ACC CCG TAC ATC
ser leu leu glu tyr leu gln gly arg ala leu his leu ala pro asn thr ala leu ala leu gly met lys thr pro tyr ile
271/91
AGC CTT CCA GTG AAA TCT ATA ATT GAT ATA CGT CAC CTT ATC GCA GGT AAA GGG CGA TCA CAC AAG CTA CCG ATA CTG ATC GAC ATG GCC
ser leu pro val lys ser ile ile asp ile arg his leu ile ala gly lys gly arg ser his lys leu pro ile leu ile asp met ala
361/121
AGC CAG ATT GCC TCG GGC ATG GCT ATC TTG AGC AGC CAG AAC TAT ATC CAC CGC GAC TTG GCC CGG
ser gln ile ala ser gly met ala ile leu ser gln asn tyr ile his arg asp leu ala ala arg
61/21
31/11
121/41
211/71
301/101
391/131
421/141

(ii)

DER	PVATKELKSTG	EELREAYIMASVEHVNLKLLAVCMSS	QMMLITQLMPLGCLLDYV	GSKALLNWSTQIAKGMSYLEEKRLVHRDLAAR
TER-a	PVAIKILKEGTA	KEFLEEAYIMASVDHPNLLKLLAVCMTS	QIMLVITQIMPLGCLLDYV	GSKPLLNWCTQIARGMAYLEDKRMVHRDLAAR
TER-b	PVAIKVLREGTQ	KEFLEEAYIMASVDHPNVLKLLAVCMTS	QIMLVITQIMPLGCLLDYV	GSKPLLNWCTQIARGMMYLEEKRMVHRDLAIR
DmSrc4/1	PVAIKTLKSGTM	KDFLAEAAQIMKKLRHTKLIQLYAVCTVE	PIYIITELMKHGSLLLEYL	SLKMQTDMAAQIAAGMAYLESQNIYIHRDLAAR
TuSrc	PVAIKTLKPGTM	KDFLEEAQATMKKLRLHPKVLQLYAVCTLE	PIYIITELMKNGSLLLEYL	KLPILIDMASQIASGMAILLSSQNIYIHRDLAAR
	*****	-*-*-*-*-*	-*-*-*-*-*	-*-*-*-*-*

Appendix 3 Figure 3.1a

(i) Nucleotides (429bp) and amino acid translation data (143aa) for *Tu Src* obtained during degenerate PCR screen for *Tu-EGFR* (c.f. Chapter 5). Grey: degenerate primers F2 + R2, with which *Tu-Src* clone was amplified. (ii) Comparative sequence alignment (90aa) of homologous domain with *Drosophila melanogaster* and *Tetrahymena urticae* putative EGFR and Src proteins. Asterisks: 100% conserved residues (36/90 = 40% shared identity). Grey: N' and C' domains corresponding to F2 and R2 primer target regions marked in (i).

orthologs are compared. An initial visual analysis based on these shared identities places *Tu*-Src firmly within the Src orthology group.

***Tu*-Src phylogenetic analysis**

Bayesian analysis of nucleotide sequences coding for the data aligned in Figure 5.4.1c produced the topology shown in Figure 5.4.1d. *Tetranychus* Src branches clearly with arthropods within a larger Src clade. *Caenorabbditis* *Let25* and *Hydra* TK genes formed notably long branches, their elevated relative rate of molecular evolution perhaps explaining points of incomplete resolution or statistical support within the tree. Alternatively, there may simply be insufficient phylogenetic signal among the 270 characters tested in the analysis; higher statistical support was obtained for the same topology generated from larger datasets (c.f. sections 5.4.2 and 5.4.3), giving weight to this latter option, as does the retention of a similar topology when Bayesian analysis is repeated without long branch taxa.

3.2 A *Tetranychus* *Tho* complex (*Tho*-C) gene³⁵

Following excision and sequencing of putative *Tu*-EGFR inserts in pBK-CMV vector, one anomalous clone was identified by BLAST as pertaining not to *EGFR* but to the *Tho* Complex (*Tho*-C) gene family. *Tho*-C genes are proposed to be related to RTK genes such as EGFR and hence could have been erroneously detected by the *TER-a* probe when screening cDNA library phage.

***Tu*-Tho-C gene sequence analysis**

Sequence data recovered for the putative *Tu*-Tho-C gene includes a 447bp open reading frame coding for a 148 amino acid polypeptide with 78bp and 59bp of 5' and 3' UTR respectively (Figure 3.2a). I constructed a multiple alignment comprising 240 amino acid characters for 30 taxa (Figure 3.2b), with Src/Abl proteins present as an EGFR outgroup, and Tho-C proteins aligned to compare proposed Tu-Tho-C domains with those of other Tho-C orthologs and TK proteins in general. Visual inspection suggests that 89 amino acids N' to the *Tu*-Tho-C stop codon align from the GGVW (G--G-V----W) conserved TK motif region (remaining *Tu*-Tho-C N' terminal 59 amino acids fall outside the alignment) (Greenwald 1985; Hsuan et al. 1989; Jorissen et al. 2003).

***Tu*-Tho-C phylogenetic analysis**

A consensus tree topology resulting from Bayesian inference analysis of aligned *EGFR*, *Src*, *Abl* and *Tho*-C orthologs shows that the *Tetranychus* *Tho*-C gene is indeed a *Tho*-C ortholog (Figure 3.2c). However, extreme divergence of *Tho*-C orthologs relative to *Src* family genes weakens support values

³⁵ c.f. Chapter V section 5.4.2

(i)

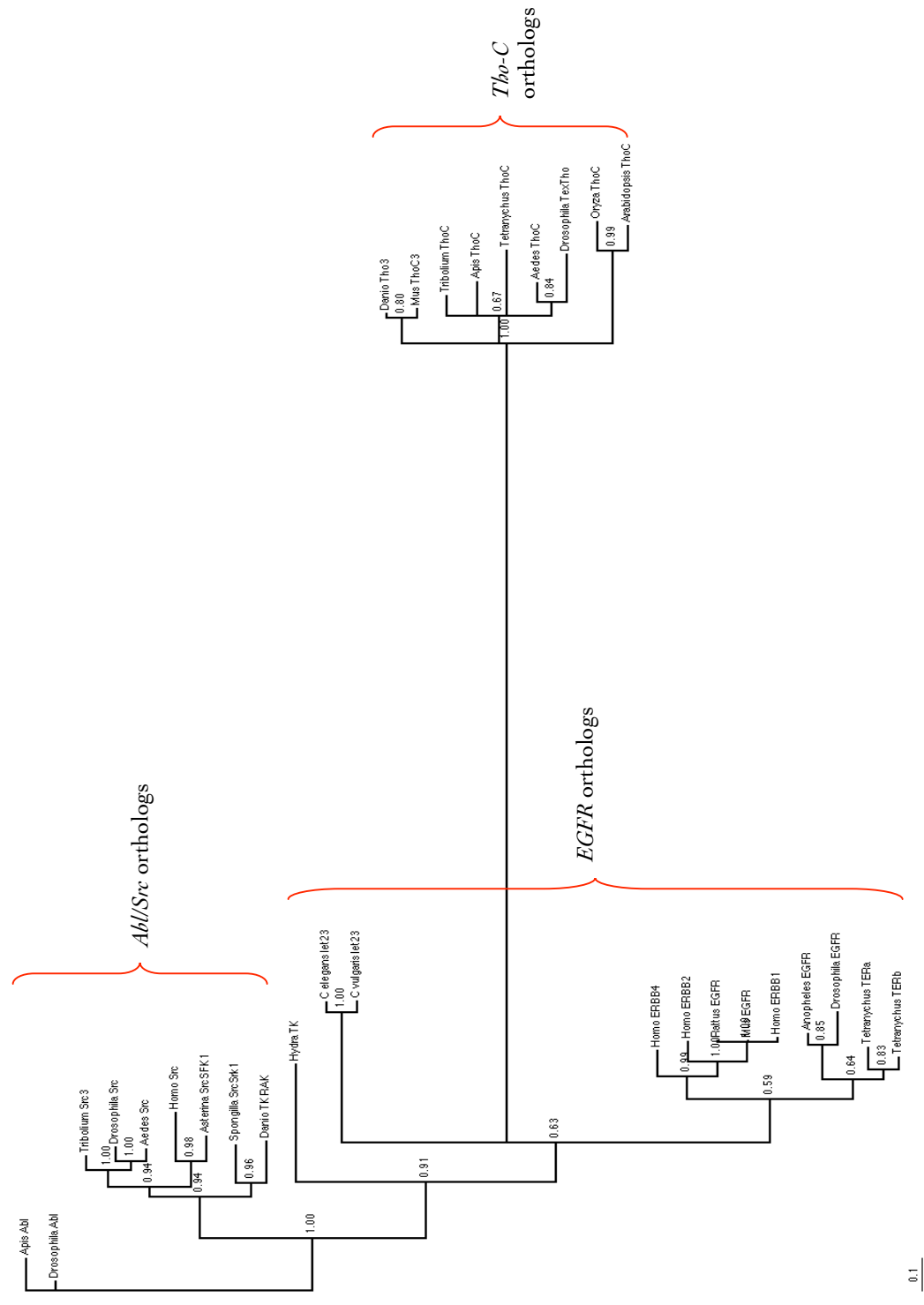
```
1          31
CAC GAG GAC CAC ATC TCG ATT TAA ACT ATT TAC ATG CTT TAT TTG TTT GTT AAA TTC AAT
61          91/5
TAA ATT ACT TTA TTT ATC ATG ACC GTG CCA TTT ATT GAA CAA TCA AAA GAA CAT TTT TCC
          met thr val pro phe ile glu gln ser lys glu his phe ser
121/15      151/25
AAA AAC AAT CGA ATC CGA GAT TAC CAA TCG CAC ACA TGT AAA GTT CAT TCT GTC GAC TGG
lys asn asn arg ile arg asp tyr gln ser his thr cys lys val his ser val asp trp
181/35      211/45
AGC TGT GAT GGA CAT CGC TTG GCT TCT GGC AGT GTA GAT AAA TCT ATT TCC ATT TTC AAT
ser cys asp gly his arg leu ala ser gly ser val asp lys ser ile ser ile phe asn
241/55      271/65
TTG GAC AAA GAT CGG TTG AAC AAG GAG AGC ACA TTG AAA GGT CAT TCA GAC TCA GTT GAT
leu asp lys asp arg leu asn lys glu ser thr leu lys gly his ser asp ser val asp
301/75      331/85
CAA CTT AGC TGG CAT CCA TCA CAT CCA GAT TTA CTG GCT ACT GCC TCA CTG GAT AAA ACT
gln leu ser trp his pro ser his pro asp leu leu ala thr ala ser leu asp lys thr
361/95      391/105
GTT CGT CTA TGG GAC TCA CGG GTT GCC AAA AAT TAC GCT ACA ATC AAT ACT AAA GGC GAA
val arg leu trp asp ser arg val ala lys asn tyr ala thr ile asn thr lys gly glu
421/115     451/125
AAC ATA AAC ATA TGT TGG TCA CCC GAT GGT TCG TCA ATT GCT GTT GGT AAC AAA GAG GAC
asn ile asn ile cys trp ser pro asp gly ser ser ile ala val gly asn lys glu asp
481/135     511/145
CTT ATC ACT TTT ATT GAC GCT AAG CAG CCA AAA AAT CAA AGT TGA ACA GCC ATT CAA ATT
leu ile thr phe ile asp ala lys gln pro lys asn gln ser -⊖-
541          571
CGA AGT CAA TGA AAT TAG TTG GAA TAA TGA CAA TGA CTT ATT CCT
```

(ii)

```
1          11          21          31          41          51          61
MTVPFIEQSK EHFSKNNRIR DYQSHTCKVH SVDWSCDGHK LASGSVDKSI SIFNLDKDRL NKESTLKGHS
71          81          91          101         111         121         131
DSVDQLSWHP SHPDLLATAS LDKTVRLWDS RVAKNYATIN TKGENINICW SPDGSSIAVG NKEDLITFID
141
AKQPKNQSO
```

Appendix 3 Figure 3.2a

- (i) *Tu-Tho-Complex* gene sequence data obtained during cDNA library screening with *TER-a* probe. CDS for 148 amino acid polypeptide, with 78bp 5' and 60bp 3' presumptive UTR sequence. Red: methionine transcription initiation codon. ⊖: TGA termination codon.
- (ii) Protein sequence summary.



Appendix 3 Figure 3.2c Bayesian consensus tree after analysis of 240 coding nucleotides from *Tu-EGFR* genes and a *Tu-Tho-complex* gene recovered during cDNA library screening for *TER* (c.f. Chapter V). Topology supports independent clades of *Abl/Src*, *EGFR* and *Tho-C* orthologs. *Tu-Tho-C* is fully supported (1.0 posterior probability) as a *Tho-C* gene, and *TER-a* and *TER-b* both group with other EGFR homologs; however, but the relationship of *Tho-C* genes is obscured by extremely long branch lengths.

and compromises branch resolution within neighbouring clades; for example, the inclusion of *Tho-C* genes disrupts an otherwise monophyletic *EGFR* homology group.

3.3 *Tetranychus Sex combs reduced*³⁶

While carrying out inverse PCR experiments, in addition to obtaining further sequence data for *Tetranychus Ultrabithorax* I also obtained by further data for the spider mite ortholog of *Sex-combs reduced* (*Scr*). PCR conditions were as mentioned in section 6.1.2 for *Ubx* iPCR experiments, except that *Tu-Scr* was amplified using three pairs of outward facing *Dfd/Scr* primers (Appendix 2.7).

Tu-Scr sequence analysis

In total, 549bp *Tu-Scr* was recovered, including 285bp open reading frame, incomplete at the 5' end, and 264bp 3' UTR (Figure 3.3a). A comparison with *Tetranychus Ubx* (Figure 3.3a) shows that while homeodomain (HD) residues are largely conserved, the two spider mite Hox proteins *Scr* and *Ubx* diverge widely beyond the 3' end of this conserved DNA binding domain. As well as identification of the extended sequence as a spider mite *Scr* homolog by BLAST identity searching, inclusion of the extended *Tu-Scr* sequence during construction of a multiple Hox alignment analysing *Tu-Ubx* (c.f. Figure 6.1.2b) permitted conclusive identification of the gene as a true *Scr* ortholog. This conclusion was supported by the presence of appropriate signature peptide motifs, and also because the resultant Bayesian consensus tree topology (c.f. Figure 6.1.2c) contains a clear *Scr* orthology group clade (de Rosa et al. 1999).

Tu-Scr gene expression?

I did not have time to synthesise *Tu-Scr* ssRNA probes to detect mRNA transcripts *in situ* in *Tetranychus* embryos. However, given the very conserved nature of collinear *Hox* complex gene expression along the antero-posterior axis in other arthropods - including chelicerates such as the spider *Cupiennius salei* and oribatid mite *Archegozetes longesitosus*, I predict that *Tetranychus Scr* is likely to be expressed in homologous body segments as those in which *Scr* orthologs are detected in other species (Damen et al. 1998; Telford and Thomas 1998a). *Archegozetes longesitosus* is the species most closely-related to *Tetranychus* for which *Scr* gene expression has been described (Telford and Thomas 1998a). *Al-Scr* is detected strongly in the third walking leg, and more weakly in the second (Figure 3.3a), so my hypothesis would be similar posterior prosomal restriction of *Tu-Scr* transcription in the spider mite.

³⁶ c.f. Chapter VI, section 6.1.2

(i) $Tu-Scr$:

[illegible]

(iii)

	1	11	21	31	41	51	61	72	81	91
TuScr	TRYQTLELEK	EFHYNRYLTR	RRRIEIAHSL	CLSERQIKIW	FONRRMKWKK	EIKFQHGSSS	YASINVRTRV	ASSSTPSPS	WRNKSIIILL	LLLL-----/
TuBx	TRYQTLELEK	EFHTSQYLTR	RRRIELAHAL	CLSERQIKIW	FONRRMKQKK	EIIQAIKELNA	QSKSSQSTSS	SSSSNNQSSN	TNSSRNHRSG	SGIT<9aa>

Appendix 3 Figure 3.3a Nucleotide and amino acid translation data for *Tu-Scr* sequence recovered by inverse PCR with Dfd-Scr primers. **Bold:** homeodomain. Underlined: previous sequence obtained by degenerate PCR. **Red:** stop codon. (i) Partial CDS and 3'UTR for *Tu-Scr*. (ii) Comparison of Tu-Ubx and Tu-Scr proteins, showing high sequence divergence (differences in grey) outside the homeodomain.

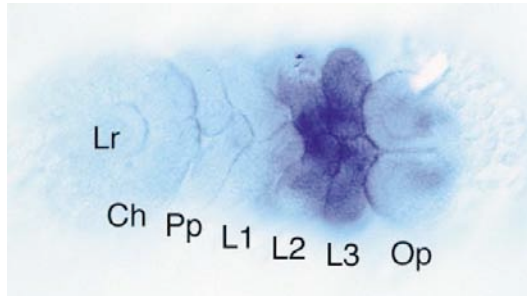


Figure 3.3a *Archegozetes longesitatus* *Scr* mRNA transcription pattern
(Telford & Thomas, 1998)

Appendix 4: Notes on phylogenetic analysis protocols

4.1 TranslatorX3 commands³⁷

TranslatorX3 is a molecular phylogenetics program that aligns in-frame coding nucleotides, processing an inferred sequence of translated codons to create an alignment within either the program Clustal-X, Muscle or T-Coffee(Jeanmougin et al. 1998). In order to run files within TranslatorX3, commands are made *via* the Terminal interface as follows:

1. `cd (TranslatorX3)`
2. `chmod +x TranslatorX3.pl`
3. `./TranslatorX3.pl`
4. `enter filename.NBRF`
5. `option 1` ('Start with unaligned nucleotides and no amino acid file')
6. `define output filename`
7. `option Y` (sequences share the same genetic code)
8. `option 1` (sequences use the Universal genetic code)
9. `C or T or M` (alignment program Clustal, T-coffee or Muscle)

³⁷ c.f. Chapter VIII, section 8.3.3

where: `cd` = change directory
(folder) = transfer named folder to Terminal shell
`./item` = find item (file/folder/program), refer back to previous location

4.2 Protocol for defining Exclusion and Inclusion Sets³⁸

1. Open FASTA format alignment in MacClade, choosing FASTA[DNA/RNA] or FASTA[PROTEIN] format as appropriate.
2. *For nucleotide alignments only:* Select all characters; within Character options, calculate 'Codon Positions' and set the position of the 1st selected base to 1.
3. *For nucleotide alignments only:* Within Display options; Colour cells of the matrix 'by translated AA state'.
4. Within Display options; Shade 'Excluded' character sets
Lighten 'Shaded' character sets
5. Adjust alignments by eye where improvements can evidently be made, using relevant tools within the program to move whole sequence strings, individual bases, or individual codons as necessary.
6. Define ExSet values; choose reliably homologous regions of the alignment (representing the InSet), and select the complementary regions (representing the ExSet, *not* to be included); MacClade specifically shades excluded cells and generates exclusion/inclusion character set values from this data.
8. Save the modified alignment and associated ExSet data as a NEXUS file.

4.3 Bayesian analysis protocol and commands³⁹

NEXUS files containing nucleotide alignments for phylogenetic analysis were subjected to Bayesian inference analysis, using MrBayes version 3.1.1 (Ronquist and Huesenbeck 2005). Expanding on the steps given in section 8.3.5 of Materials and Methods Chapter VIII, more full details of steps used in the processing of each file to run Bayesian inference analysis are given in the following points.

1. Open .nex alignment in BBEdit, PAUP or TextWrangler to view the aligned matrix of taxa and character, codon and ExSet information in standard text format.

³⁸ c.f. Chapter VIII, section 8.3.4

³⁹ c.f. Chapter VIII, section 8.3.5

2. Write 'MrBayes Block' beneath the matrix, defining a set of instructions about characters for the program to include. An example Bayes Block is:

```
BEGIN mrbayes;  
CHARSET 3rdposn = 3-1749\3;  
CHARSET exclude = 1-576 820-1749;  
exclude exclude;  
exclude 3rdposn;  
lset nst = 6 rates = gamma ngammacat = 4;  
mcmc ngen = 1000000 samplefreq = 100 printfreq = 100  
END;
```

Within the Bayes Block, notable features are:

- (i) Third codon positional values (3rdposn), taken from within the CODONPOSSET/BEGIN SET data in the nexus text file. Third codon positions are excluded due to high functional lability at the third base of any given codon and the likely associated 'saturation' as multiple transformations and reversions occur, masking useful phylogenetic information.
 - (ii) Exclusion set values, given in the nexus text file, present as EXSET values nested within Bayes Block BEGIN SET/BEGIN ASSUMPTIONS.
 - (iii) The number of 'rates', referring to complexity of assumptions made in the phylogenetic model. For example, two states treats transitions and transversions as being of unequal probability and evolutionary significance relative to the nature of the 'ancestral' base, whereas consideration of 6 states can distinguish more subtle factors affecting the probability of a given base changing/reverting to any other.
 - (iv) At line 7, ngen refers to 'number of generations', in the form of searches made for the 'best tree', and samplefreq ('sample frequency') refers to the interval between recordings, or frequency at which tree/Log Ln values are recorded whilst the analysis is running.
3. So that the NEXUS file can be processed correctly in MrBayes, the following modifications were made to the input data within a text program:
- delete all superfluous 'comment' brackets (square brackets) and their contents, preventing them from being mis-read as instructions or characters in the phylogeny program
 - delete all command blocks other than the Mr Bayes Block, such as raw character set data and commands used by alignment programs
 - check necessary gaps and spaces present in Bayes Block commands (e.g. no gap in '3rdposn')
 - convert and save file in Unix format, most easily read by Terminal
4. Once the file is formatted with the relevant Bayes Block, it is placed in the MrBayes folder and a new shell opened in Terminal to execute the Bayesian inference program with the following commands:

cd (Mr Bayes) = change directory to 'Mr Bayes' (drag folder onto shell)
 ./mb = go to phylogeny program within Mr Bayes folder
 exe file.nex = execute analysis of nexus file named in Mr Bayes folder

5. Once the number of specified generations have been analysed, various output files are created, relevant ones being suffixed '.t' and '.p'. Files suffixed .t contain tree morphology information, and .p (probability) files the log likelihood (LnL) data for each tree permutation recorded at each generation in the search for the 'best tree', required to assess stability of the consensus tree.

Appendix 5: *Engrailed* gene multi-functionality

Engrailed transcription factors are widely multi-functional, with En-mediated signalling implicated as particularly important to both segmentation and neurogenesis throughout arthropods as well as in more distant phyla. This being so, the complete absence of *Tetranychus urticae en-1* and *en-2 in situ* hybridisation signals in all embryos, from late cleavage to late organogenesis, is very unlikely to be due to genuine absence of transcriptional activity. Major common themes in *Engrailed* gene function in metazoans are detailed in the following sections:

5.1 *Engrailed* and segment boundary formation

Engrailed gene products are known to define anterior parasegment boundaries and maintain the posterior fate of segmental structures, including appendages, throughout the arthropods. This critical function has been proved by loss-of-function experiments, for example classical genetics in *Drosophila* and RNAi in *Oncopeltus fasciatus*, as well as gene expression studies in species from all major arthropod phyla (Akam 1987; Angelini and Kaufman 2005a; Damen 2002; Gibert et al. 2000; Hughes and Kaufman 2002a; Karr et al. 1989; Prpic 2004a). A segmental boundary role for *engrailed* appears to exist outside the arthropods - e.g. polychaete annelids *Platynereis* and *Chaetopterus*, glossiphoniid annelid *Helobdella triserialis* (Lophotrochozoa), and in primary *Amphioxus* somites (Deuterostomia: Echinodermata). This suggests a possible conserved role in segment boundary formation since divergence of the protostome-deuterostome common ancestor (Prud'homme et al. 2003; Scholtz 2001; Wedeen et al. 1997; Wedeen and Weisblat 1991). Although the homology of this boundary-

defining role outside the arthropods may be disputable, as a chelicerate *Tetranychus urticae* would be strongly predicted to exhibit some form of segmental *engrailed* expression, even if one or both of its paralogs did not report complete transverse stripes or continuous temporal and spatial expression patterns, as is the case for *Cupiennius en-1* and *en-2*, and to a lesser degree in a number of crustaceans and insects with two *engrailed* paralogues (e.g. *Sacculina*, *Porcellio*, *Procambarus*, *Orchestia*, *Thermobia*)(Damen, 2002; Quéinnec, 1999; Gibert, 2000; Gibert, 2002; Abzhanov, 2000; Hejnol, 2004; Peterson, 1998).

5.2 *Engrailed* and segmental limb outgrowth

Limb outgrowth in all arthropods examined commences just anterior to sharply defined *engrailed* stripes, indicating a conserved signalling interaction at the parasegment boundary that may instruct initial limb outgrowth. In the cellularising *Drosophila* embryo, having been established by reiterated pair-rule gene interactions, *engrailed* is activated in 14 single-cell wide transverse stripes (in A-P and D-V progression); activation of *wingless* in cells anterior to the *engrailed* row creates an A-P parasegment boundary organising centre that is strictly maintained by an auto-regulatory positive feedback loop between Hh (activated by En) and Wg morphogens, conferring a sharp anterior limit to *engrailed* stripes. Once parasegments are established, *en* and *wg* are expressed independently of each other (Akam 1987; Alexandre and Vincent 2003; Deutsch 2004; DiNardo et al. 1985; Ingham and Martinez Arias 1992; Jaynes and Fujioka 2004). It is tempting, given the apparent conservation of the arthropod parasegment, to posit *Drosophila*-like Engrailed-Wingless interactions during early induction of *Distal-less* expression and ectodermal limb outgrowth, with Wg acting as an activator of *Dll* transcription at the intersection of A-P and D-V boundaries. Although numerous expression studies in diverse arthropod species would support such a hypothesis(Damen 2002; Hughes and Kaufman 2002a; Prpic 2004a; Scholtz 1997; Scholtz 2001), recent studies in certain crustaceans and insects call into question an absolute requirement for En/Wg to activate *Dll*, even when temporal and spatial gene expression patterns suggest otherwise. For example, in the malacostracan crustacean *Orchestia cavimana*, *Dll* is first detected before *engrailed* is expressed, in *Mysidium columbiae* *wingless* expression is not associated with segments at all, and in the insects *Gryllus bimaculatus* and *Oncopeltus fasciatus*, RNAi only disrupts limb development when *engrailed*, but not *wingless*, transcription is abrogated, in spite of adjacent striped expression of both genes(Angelini and Kaufman 2005a; Duman-Scheel et al. 2002; Hejnol and Scholtz 2004; Miyawaki et al. 2004).

5.3 *Engrailed* orthologs in neuro-, myo- and morpho-genesis

Further to determining the posterior fate of segmental structures in arthropods, *engrailed* expression reveals other possible conserved functions. A conserved role in nervous system patterning throughout arthropods and the wider metazoa is supported by *en* activity in the arthropod CNS and/or PNS, onychophoran CNS, annelid CNS, chordate neural tube and vertebrate brain (Abzhanov and Kaufman 2000b; DiNardo et al. 1985; Gibert et al. 2000; Hejnol and Scholtz 2004; Hemmati-Brivanlou et al. 1991; Patel et al. 1989; Peel et al. 2006; Prud'homme et al. 2003; Wedeen et al. 1997; Wedeen and Weisblat 1991). Coupled to the ability of En-mediated signalling to direct changes in cellular adhesion and form, *en* expression is associated with segmental groove morphogenesis and myogenesis in arthropods and onychophora, heart formation in Pancrustacea, and morphogenesis of the eye and various posterior structures (e.g. terminalia, caudal cerci, proctodeum, hindgut) in a range of arthropods (Chipman et al. 2004a; Chipman et al. 2004b; Damen 2002; Larsen et al. 2003; Peterson et al. 1998; Simonnet et al. 2004; Wedeen et al. 1997).

5.4 Ancestral functions of arthropod *Engrailed* orthologs?

On the basis of comparative *engrailed* expression patterns in the annelids *Helobdella* and *Platynereis*, onychophoran *Akanthokaris* and many arthropods (also considering *Euscorpius* *hh* expression, activated directly by Engrailed), some hypotheses about ancestral roles of engrailed can be posited: In the common ancestor of annelids and arthropods, i.e. a basal protostome, *engrailed* may have had a double role in 1. defining segment polarity to the ectoderm/mesoderm (perhaps connected to initial ganglion localisation) and 2. defining segmental ganglion fate during neurogenesis. Added to these two roles, in the common ancestor of arthropods there appears a third, in patterning terminal elements of the digestive tract (Davis and Patel 2002; Prud'homme et al. 2003; Simonnet et al. 2004; Wedeen et al. 1997; Wedeen and Weisblat 1991).

Appendix 6: Detection of *Tu-Sp8/9* and *Tu-Sp1/3/4* transcripts with short ssRNA probes

6.1 Single stranded RNA probe synthesis and efficacy

i) Probe synthesis

In spite of the minimal size of gene fragments so far cloned, ssRNA DIG-AP labelled probes were synthesised from purified pGEM®-T Easy plasmids carrying ~170bp *Tetranychus Sp* gene inserts, to test for any ability to detect mRNA expression *in situ*. EcoRI restriction enzyme digestion allowed confirmation of correct insert size prior to probe synthesis (Figure 6.1,i-ii), and as all the relevant clones were inserted in 5' → 3' orientation, T7 and Sp6 RNA polymerases were used for sense(+) and anti-sense(-) strand synthesis respectively (Figure 6.1,i).

ii) DIG-labelling efficiency

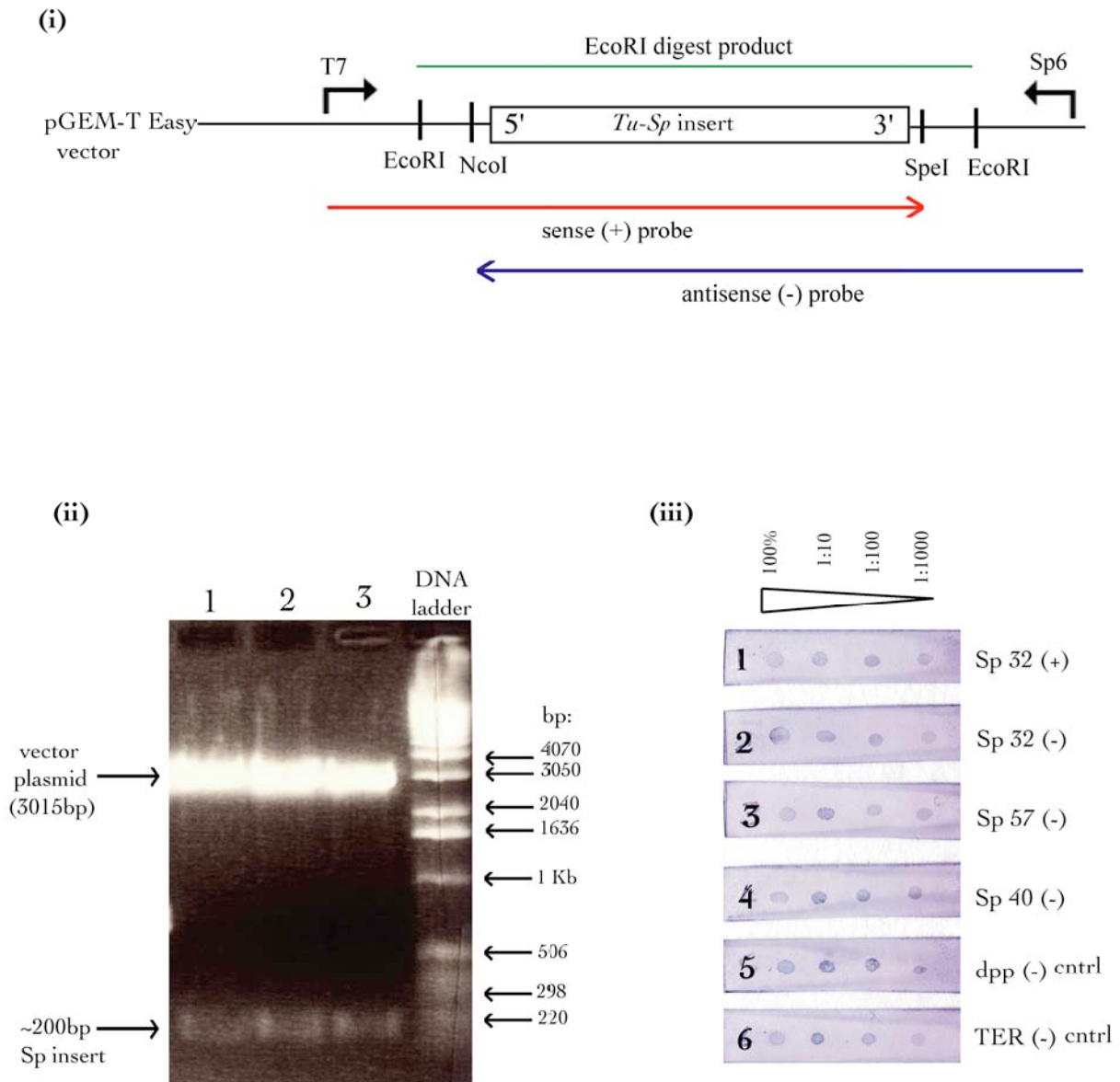
Efficient DIG labelling was confirmed by hybridisation with α-DIG-AP (Figure 6.1,iii). The “Sp 40(-)” probe against *Tu Sp1/3/4* gave consistently greater signal intensity than the “Sp 32(+)/(-)” and “Sp 57(-)” probes against the *Tu Sp8/9* ortholog. This signal discrepancy suggests slightly higher labelling efficiency in the probe against *Tu Sp1/3/4*, although probes against *Tu Sp8/9* were clearly adequately labelled to produce a positive signal in an AP-H₂O₂ reaction. However, efficient binding of such short probes to their target within the transcriptome is a much less certain factor, and shall be discussed with reference to problems encountered in obtaining whole mount *in situ* hybridisation results.

6.2 Expression of *Tu-Sp1/3/4* and *Sp8/9* during embryogenesis

6.2.1 *Tu-Sp8/9* mRNA transcription

i) Negative results

I detected tentative positive expression in blastoderm stage embryos (7-9hr AEL), in a broad, unilateral domain (see paragraph ii) below), but no other *Sp8/9* was detectable at any later stage, calling the validity of the former into question. The only available *Sp8/9* RNA probe was ~180bp in size, arguably several hundred bases shorter than would be recommended for effective probe-target



Appendix Figure 6.1

(i) Schematic diagram of a putative *Tu-Sp* gene insert, cloned into pGEM-T Easy vector plasmid, with elements related to insert size verification and synthesis of ssRNA probes. An EcoRI digest of purified plasmid yields an 'insert + ~20bp vector' fragment (green line).

(ii) 1% agarose gel, lanes 1-3 containing excised *Sp* inserts (~200bp arrow). Vector plasmid (3015bp) linearised by EcoRI digestion, runs as the high molecular weight upper band.

(iii) Results of hybridisation with an anti-DIG-AP antibody to test for effective DIG-labelling of RNA probes. Serial dilutions were made in DEPC water, from RNA stock in HYB buffer (hence the 100% dot is less clear), and DIG-labelled mRNA cross-linked to nitrocellulose strips prior to hybridisation. All probes give a typical pattern of fading phosphatase reaction signal with increasing dilution factor.

recognition. In order to discern whether lack of evidence for *Sp8/9* mRNA is explained by genuine transcriptional inactivity or by sub-optimal probe size, extension of available sequence data for *Tu Sp8/9* is required. More complete cloning would allow synthesis and testing of improved ssRNA probes *in situ*. Other possible reasons for failure of *Sp* probes to detect mRNA transcripts include probe RNA degradation or mRNA disruption during fixation. Probe RNA may theoretically have degraded, although every precaution against RNases was taken, but it seems unlikely that embryonic mRNA degradation caused the failure, as embryos fixed with the same protocol retained detectable amounts of other mRNA transcripts (e.g. *Dll* antisense control, Figure 6.2,i).

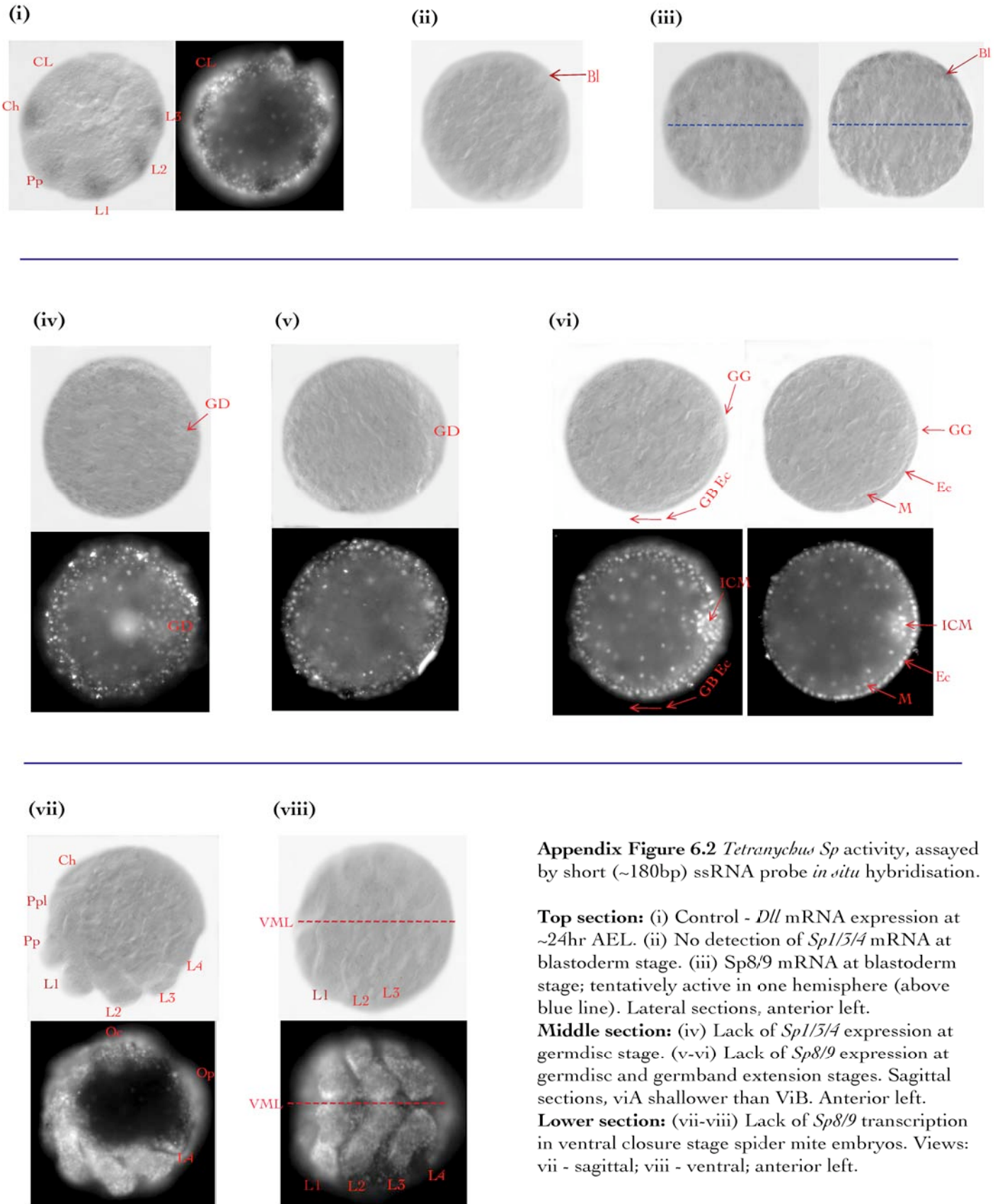
ii) Possible blastoderm stage expression?

I observed an approximately hemispherical pattern of weak but consistent *Tu Sp8/9* mRNA expression in early (7-9hr AEL) blastoderm embryos (Figure 6.2,iii). The orientation of the domain cannot be discerned and hence proposals as to its plausible fate or function are limited, discussed as follows:

Arthropods: Sp function along the antero-posterior axis

In *Drosophila*, *Sp8* and its duplicate *buttonhead* are expressed early in a broad anterior stripe in the syncytial blastoderm, traversing future antennal, intercalary and mandibular segments. Upstream control of *bt* expression is achieved via integration of multiple maternal signals, each providing spatial co-ordinates for activation or repression (Shandala et al. 1999; Wimmer et al. 1996; Wimmer et al. 1995). Alone, this anterior specification role could hint at anterior-posterior axial determination that may be conserved in *Tetranychus*, but a consideration of *Drosophila* embryogenesis and comparison with other arthropod studies, do not support the idea:

- i) The *Tetranychus* germband arises from a germdisc in a manner distinct from that of the long germband insect *Drosophila*, such that any ancestral axial specifying role could only be extended to the process of breaking symmetry, possibly along an emerging A/P axis, as the germband is accumulating;
- ii) Earliest *Sp8* activity in the beetle *Tribolium* is in segmental germband stripes, indicating lack of a conserved *Drosophila*-type anterior domain, even within Hexapoda (Beermann et al. 2004);
- iii) In the absence of direct *Sp* data for other chelicerates, *Notch* - an upstream activator of *Sp8* - is expressed in early segmental stripes in the spider *Cupiennius*, rather than in any A-P domain (Stollewerk et al. 2003a; Stollewerk et al. 2003b).



Appendix Figure 6.2 *Tetranychus Sp* activity, assayed by short (~180bp) ssRNA probe *in situ* hybridisation.

Top section: (i) Control - *Dll* mRNA expression at ~24hr AEL. (ii) No detection of *Sp1/5/4* mRNA at blastoderm stage. (iii) *Sp8/9* mRNA at blastoderm stage; tentatively active in one hemisphere (above blue line). Lateral sections, anterior left.

Middle section: (iv) Lack of *Sp1/5/4* expression at germdisc stage. (v-vi) Lack of *Sp8/9* expression at germdisc and germband extension stages. Sagittal sections, viA shallower than ViB. Anterior left.

Lower section: (vii-viii) Lack of *Sp8/9* transcription in ventral closure stage spider mite embryos. Views: vii - sagittal; viii - ventral; anterior left.

Bl - blastomere layer; GD - germdisc; GB Ec - germband ectoderm; GG - gastral groove; M - mesoderm; VML - ventral midline; CL - cephalic lobe; Oc - ocular lobe; Ppl - pedipalp lobe; Pp - pedipalp; Ch - chelicera; L - walking limb; Op - opisthosoma; ICM - inner/endodermal cell mass.

Vertebrates: Sp function along the dorso-ventral axis

In vertebrates, most reports of *Sp* gene roles relate to maintaining signals during earliest limb outgrowth (e.g. FGF, Wnt and BMP at the apical ectodermal ridge) and anterior neurogenesis (Bell et al. 2003; Triechel et al. 2003). However, studies in *Danio* have revealed *Sp5* and *Sp5-like* function in early dorso-ventral axis determination within mesoderm and neuroectoderm tissues (Thorpe et al. 2005; Weidinger et al. 2005). Elements of such dorso-ventral specification may be paralleled in the polarised early spider mite *Tu Sp8/9* domain, although such ideas are highly speculative and depend on further gene sequencing, more reliable *in situ* mRNA detection and data from arthropods, with whom direct genetic homology is more validly discussed.

iii) Hypotheses for *Tu-Sp8/9* arising from comparative data

As mentioned in the Introduction (section 1.2.3), *Sp8/9* function has been demonstrated in the insects *Drosophila* and *Tribolium*, including roles in segment formation, limb allocation, proximo-distal limb outgrowth (allometric or otherwise) and anterior CNS/PNS development (Beermann et al. 2004; Estella et al. 2003; Wimmer et al. 1997; Wimmer et al. 1996; Wimmer et al. 1995). Therefore, although the spider mite germdisc may use distinct signalling systems during its aggregation, and is a challenging structure to homologise with other arthropods examined so far (i.e. long germband *Drosophila* and intermediate germband *Tribolium*), as a member of Arthropoda I would predict some form of *Sp8/9* activity, however divergent, at or beyond this stage, as the segmented, appendage-bearing germband develops. Outside the arthropods, crucial roles for vertebrate *Sp8* and *Sp9* orthologs have been shown in maintaining AER signalling during limb outgrowth, both anterior and posterior neuropore closure, and posterior axial development; the latter role in concert with *Sp5* (Bell et al. 2003; Thorpe et al. 2005). Conserved *Sp8/9* gene functions in limb outgrowth and anterior CNS development in protostomes (arthropods) and deuterostomes (vertebrates) imply similar roles in an Urbilaterian ancestor, and strengthens the expectation that *Sp8/9* gene expression patterns in *Tetranychus* are absent in this case not due to lack of transcription but due to poor *in situ* hybridisation. One would expect some form of limb associated *Tu Sp8/9* expression to be observed, were mRNA transcripts properly detected.

6.2.2 *Tu-Sp1/3/4* mRNA transcription

i) Negative results

In situ hybridisation with short (~170bp) *Tu Sp1/3/4* ssRNA probes produced apparently negative results at all stages of embryonic development. Lack of demonstrable *Sp1/3/4* transcription (e.g. at

blastoderm stage, Figure 6.2,ii) may indicate true *Sp1/3/4* gene inactivity, or more likely that the ~170bp RNA probe was too short for sufficient target mRNA recognition and binding. Extension of the *Tu Sp1/3/4* sequence is required to synthesise longer probes, strengthening inter-molecular probe-target bonding and allowing better investigation of *Sp1/3/4* transcription during spider mite embryogenesis.

ii) Possible specific expression at germdisc stage?

Germdisc stage embryos assayed for *Tu Sp1/3/4* mRNA activity seem to show weak ectodermal transcription in the apparent ventral hemisphere, adjacent to and slightly overlapping the germdisc at its anterior end (Figure 6.2b,iv). This DIG-AP staining pattern is likely an artefact produced from fixation or staining conditions, or it may indicate real transcription, and tentatively a role for *Sp1/3/4* in anterior germband specification⁴⁰, reminiscent of blastodermal *Sp8/btd* in *Drosophila* head/brain development and *Sp8* function in vertebrate anterior CNS (Bell et al. 2003; Gallitano-Mendel and Finkelstein 1998; Triechel et al. 2003; Younossi-Hartenstein et al. 1997).

Appendix 7: Amended interpretations of data published on early segmentation and germ cell fate specification in *Tetranychus urticae*

7.1 Early segmental patterning

Aside from consistent scale bar errors and several instances of mis-match between quoted hrs AEL and stage of embryonic development for corresponding embryos, the most significant amendments required of results published in 'Dearðen, Donly e³ Grbic (2002) *Expression of pair-rule gene homologs in a chelicerate: early patterning of the two-spotted spider mite Tetranychus urticae*' are listed as follows:

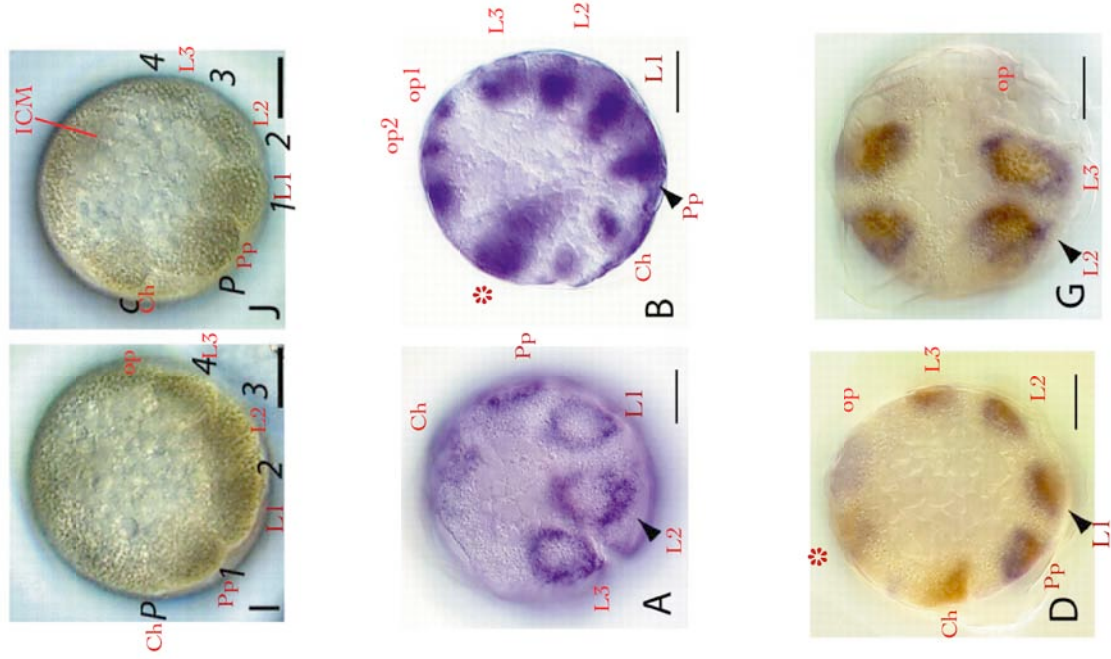
- a. In giving an overview of spider mite development (p. 5463) the authors do not attend to events surrounding dynamic morphogenesis between blastoderm and germband stages, cell

⁴⁰ More tenuously, the domain could be linked to ventral specification, reminiscent of *Danio Sp5*'s role in Wnt-mediated D/V mesoderm and neuroectoderm patterning, and the *mBtd*-dependent maintenance of D/V *En1* and *BMP4* expression during mouse limb development (Bell et al. 2003; Triechel et al. 2003; Weidinger et al. 2005).

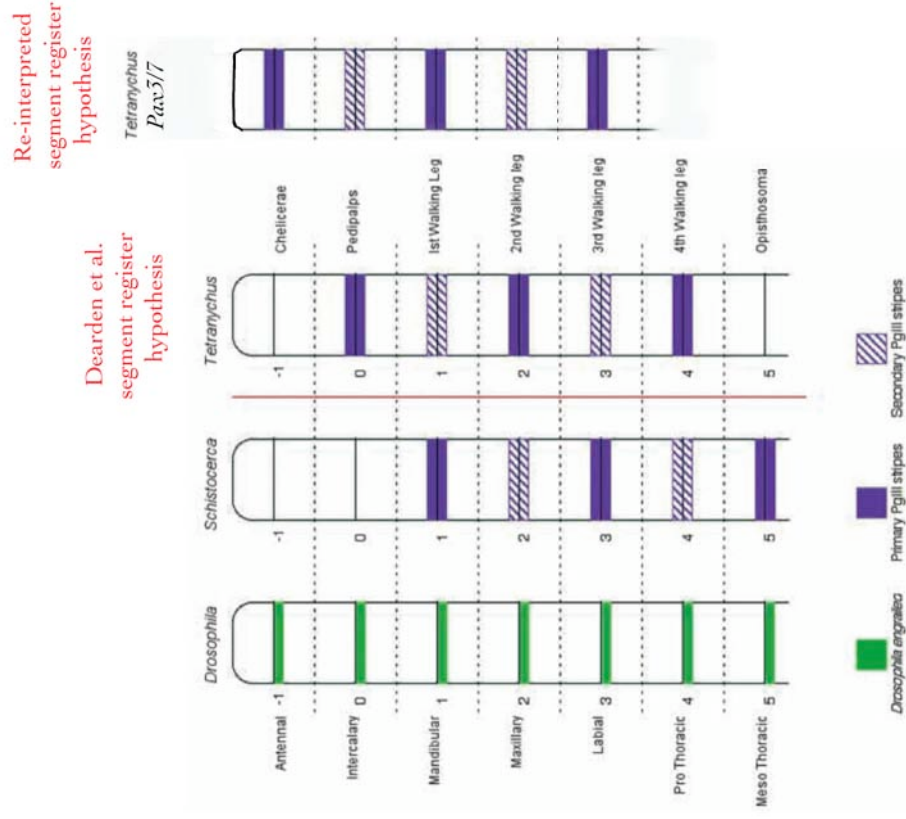
movements and germ layer differentiation during germ band development, dynamic anterior appendage morphogenesis nor the final prolonged phase between final embryogenesis and maturity of the pre-larval stage before hatching. Lack of precise observation and timing is the source of later confusion and likely mis-interpretation of cell and segment identity.

- b. Labelling of prosomal segments (and in some case the antero-posterior axis) on the basis of morphology and/or Dll protein domains appears to be wrong for certain embryos in Fig.1 (p.5463), Fig. 4 (p.5466), Fig. 5 (p. 5467) and Fig. 6 (p.5469). Most typically this arises from failure to observe that the L4 segmental appendage is not specified until late in embryogenesis, so segment labels require adjustment by usually one, if not two segments to the anterior.
- c. The report of Dll protein dynamics during embryogenesis (p. 5464) does not note the transient initial stage of 4 paired Dll domains corresponding to cheliceral segment delay, and wrongly assumes that the 5 early paired domains correspond to paired pedipalp and four walking limb primordia, whereas these in fact correspond to primordial cheliceral, pedipalp and three walking limb pairs. No notice is given to Dll activity in cephalic lobes either, which results in some confusion between the cheliceral segment and cephalic lobe when labelling segments, also related to trying to accommodate a 4th limb pair (see Figure 7.1,i).
- d. The interpretation of *Tu-runt* gene expression (Fig. 4) is based on cross-reference with incorrectly defined Dll domains, working backwards with a certain degree of arbitrary temporal inferences given that detectable mRNA would have appeared at random in a minor sample of mass-prepared embryos at mixed stages.
- e. The Pax-III group gene *Tu-Pax3/7* is stated to be expressed like a segment polarity gene regulated by pair-rule genes to form primary and secondary stripes at a segmental register one segment anterior to those of insects such as *Schistocerca gregaria* (p. 5466-5469). Unfortunately the mis-identification of segments is transferred to *Tu-Pax3/7* + Tu-Dll double-stained embryos, with the general result that primary stripes of Tu-Pax3/7 expression are linked to 'even' segments and secondary stripes to 'odd' segments (Figs 4 – 7). When segments are labelled correctly, however, this displaced segmental register returns into a conserved register with insects, and *Tu-Pax3/7* is also then excluded from expression in the anterior opisthosoma (see Figure 7.1,ii).
- f. When interpreting *Tu-Pax3/7* gene expression during later embryogenesis (p. 5467, Fig. 5), down-regulation at the ventral mid-line is inferred, whereas an appreciation of tissue movement as the ventral ridge closes may better explain the observed faded gene expression.
- g. Fragments of, or entire, vitelline membranes are visible around many of the embryos illustrated in Figures 4 – 6, calling into question the reliability of transcription patterns

(i)



(ii)



Re-interpreted
segment register
hypothesis

Dearden et al.
segment register
hypothesis

Tetranychus
Pax5/7

Appendix 7 Figure 7.1 Amendments to Dearden et al. 2002.

(i) Re-identification of appendage segments - my new identifications in RED font. Embryos I + J (DIC) from Fig.1 p.5463 in Dearden et al. Embryos A + B (*Tu-run* mRNA) from Fig.4 p.5466. Embryos D + G (*Tu-run* blue + *Tu-Dll* brown) Fig.4.

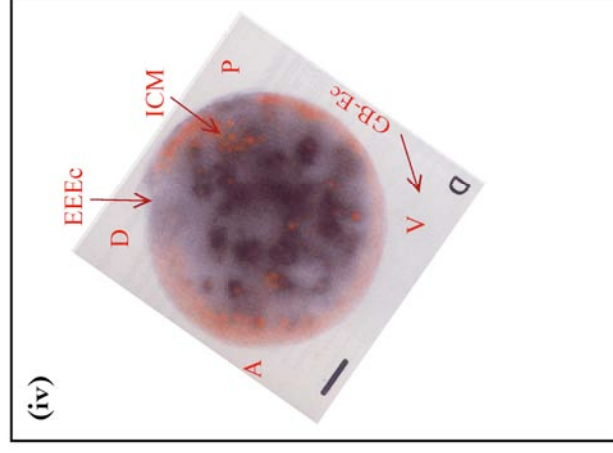
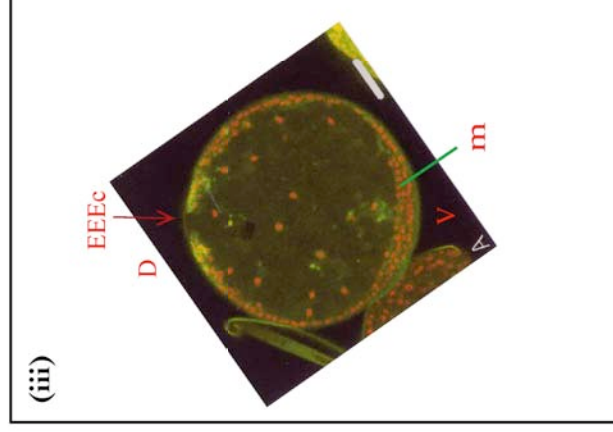
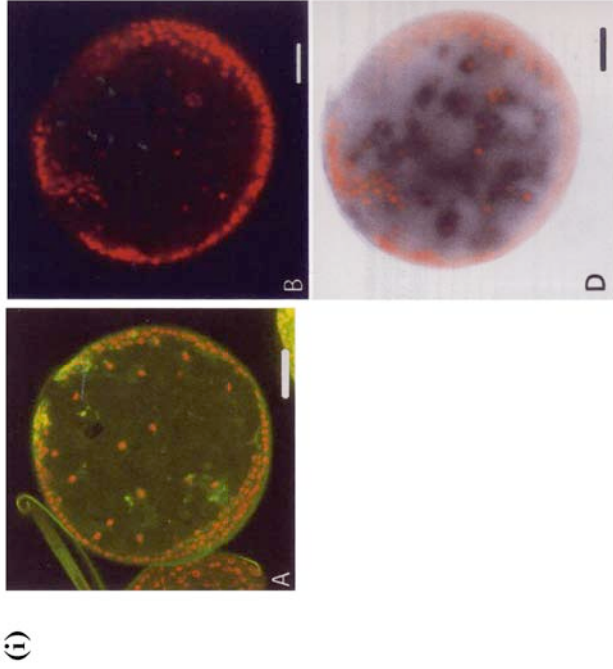
(ii) Re-interpretation of segmental register defined by Pg-III gene *Tu-Pax5/7* during early embryo patterning. *Tu-Pax5/7* stripes are in register with orthologous gene activity in *Schistocerca*, but start slightly more anteriorly r.e. segment homology.

detected beneath, as staining may be disrupted or biased by uneven or incomplete RNA probe penetration.

7.2 Germ line specification

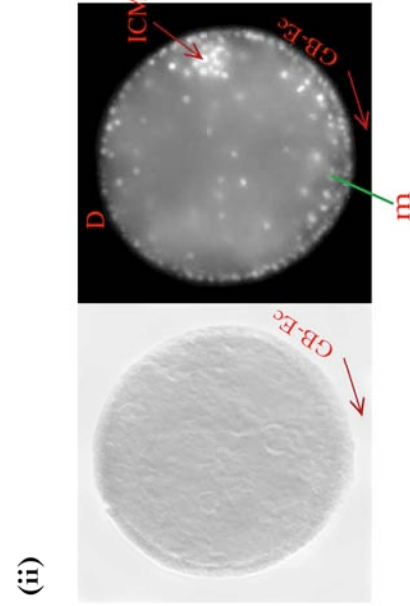
Aside from scale bar errors and vitelline membrane material on embryo preparations as in Dearden et al. (2002), the most significant comments to make on results published in 'Dearden, Grbic & Donly (2005) *Vasa expression and germ-cell specification in the spider mite Tetranychus urticae*' are as follows:

- a. There are mis-matches between quoted stage of development, hrs AEL and the actual stage of development of embryos supposed to represent blastoderms. For example, in Fig.2 (p. 601) an embryo described as a blastoderm is quoted as being at 17hr AEL, a time which refers to late germ disc or early germ band stages even according to the authors' own spider mite development time scale(Dearden et al. 2002). Of more significant consequence, the authors associate 'blastoderm stage' *vasa* gene expression in Fig.3 with intra-yolk cells, whereas such embryos are clearly not blastoderm stage but rather germband stage embryos with an evident sub-opisthosomal cell cluster for orientation and mesoderm cells visible beneath the germband ectoderm (see Figure 7.2). This correction, in addition to observed asymmetry and blurriness in detected *Tu-vasa* expression lead me to doubt the validity of assigning significance to apparent *Tu-vasa* activity. Early *Tu-vasa* detection could be an artefact related to properties of intra-yolk, vitellogenic cells, although an alternative role in mediating earliest endoderm or mid-gut cell specification may be possible, given apparent accumulation of some *Tu-vasa*-positive cells in the region of the sub-opisthosomal cell cluster in mid-later germband stages (Fig. 3).
- b. *Tu-vasa* expression is strong and distinct from ventral ridge closure stages onwards, localised to a clear endo- and/or meso-dermal domain that is suggestive of true affiliation to gonad tissue and possible specification of germ cells *after* germ layers are differentiated, unlike the hypothesis of germ cell specification from non-blastoderm cells separated from main embryo cells since early blastoderm formation.
- c. The *Tu-vasa*-positive cluster of cells occurs in a domain that the authors describe as dorsal to the L4 limb segment, but this identity is based on misplaced assumptions about the complement of specified limbs at earlier stages in *Tetranychus* embryogenesis. In fact, *Tu-vasa* cells are beneath the anterior opisthosomal region, or 'dorsal' and posterior to the L3 segment.



Appendix 7 Figure 7.2 Amendments to Figure 3 from Dearden et al. 2003.

- (i) Original Figure 3 from Dearden et al. (2003), proposing to show two blastoderm stage embryos (A and B+D).
- (ii) *Tetranychus* germband stage embryo from Chapter II Figure 2.1.3a.
- (iii) Interpretation of Fig.3A as a transverse section, germband stage.
- (iv) Interpretation of Fig.3D as sagittal section, germband stage embryo.



N.B. Scale bars proposed as 50 microns by Dearden et al. I calculate to be 16 microns in A and 14 microns in B+D.

m - mesoderm; GB-Ec - germband ectoderm; A - anterior; P - posterior; D - dorsal; V - ventral; ICM - inner, ectodermal cell mass; EEEc - extra-embryonic ectoderm